

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 1, 2003, 05:18:35 ; Search time 585.641 Seconds

(without alignments)
993.879 Million cell updates/sec

Title: US-10-053-662A-29

Perfect score: 20

Sequence: 1 tttaccacagggagatgagaa 20

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : GenEmbl1.*
1: gb_ba:*
2: gb_hcg:*
3: gb_in:*
4: gb_ov:*
5: gb_om:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_scs:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vl:*
30: em_hcg_hum:*
31: em_hcg_inv:*
32: em_hcg_other:*
33: em_hcg_mus:*
34: em_hcg_dln:*
35: em_hcg_rtd:*
36: em_hcg_mam:*
37: em_hcg_vrt:*
38: em_sy:*
39: em_hcg_hum:*
40: em_hcg_mus:*
41: em_hcg_other:*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	18.4	92.0	223	9	HS1AM110
2	18.4	92.0	2310	6	AK021648
3	18.4	92.0	3620	6	AX045367
4	18.4	92.0	3720	6	AX045565
5	18.4	92.0	4316	6	162750
6	18.4	92.0	4316	6	HS1AMB2TB
7	18.4	92.0	5020	6	AX045563
8	18.4	92.0	5156	6	AX365737
9	18.4	92.0	5156	9	HSNICE
10	18.4	92.0	5200	6	AX045561
11	18.4	92.0	5200	6	162749
12	18.4	92.0	5200	6	HS1AMB2T
13	18.4	92.0	98233	2	AC131536
14	18.4	92.0	133936	9	HS0J38304
15	18.4	92.0	164084	9	AL354953
16	18.4	92.0	170492	2	AL139013
17	17.4	87.0	31720	9	HS1196E3
18	17.4	87.0	80327	2	AC055889
19	17.4	87.0	145917	2	AC098602
20	17.4	87.0	146576	8	AC074282
21	17.4	87.0	159382	8	AC116601
22	17.4	87.0	164682	9	AC006199
23	17.4	87.0	165080	2	AC117493
24	17.4	87.0	182843	2	AC095080
25	17.4	87.0	196476	2	AC055890
26	17.4	87.0	199255	2	AC084195
27	17.4	85.0	175288	2	AC094555
28	16.8	84.0	63892	2	AL450339
29	16.8	84.0	64719	2	AC129496
30	16.8	84.0	67406	2	AC099897
31	16.8	84.0	74289	9	AC011813
32	16.8	84.0	84234	9	AC107975
33	16.8	84.0	94531	2	AC106626
34	16.8	84.0	100000	9	AB020875
35	16.8	84.0	120978	9	AL732367
36	16.8	84.0	123071	2	AC121481
37	16.8	84.0	123706	3	TBR25N7
38	16.8	84.0	136788	2	AC117044
39	16.8	84.0	144290	9	AC098866
40	16.8	84.0	146340	9	AP003479
41	16.8	84.0	148159	2	AC106365
42	16.8	84.0	151453	2	AC118988
43	16.8	84.0	155655	2	AC118843
44	16.8	84.0	153808	2	AC105845
45	16.8	84.0	158561	2	AC114125

ALIGNMENTS

RESULT 1
LOCUS HS1AM110
DEFINITION Human laminin gamma2 chain gene (LINC2), exon 10 and flanking sequences.
ACCESSION U31187.1 GI:1236307
VERSION U31187.1
KEYWORDS
SEGMENT
SOURCE
ORGANISM Homo sapiens.
Homo sapiens.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (sites)
AUTHORS Kallunki,P., Sainio,K., Eddy,R., Byers,M., Kallunki,T., Sariola,H.,

ADVKNLENIRDNLPFGCYNTQALEQO"
BASE COUNT 959 a 862 c 1040 g 759 t
ORIGIN

Query Match 92.0%; Score 18.4; DB 6; Length 3620;
Best Local Similarity 95.0%; Pred. No. 29;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 TGTACTCAGGGATGAGAA 20
|||||
Db 1228 TGTATTCAGGGATGAGAA 1247

RESULT 4
LOCUS AX045565 3720 bp DNA linear PAT 24-NOV-2000
DEFINITION Sequence 29 from Patent WO0066731.
ACCESSION AX045565
VERSION AX045565.1 GI:11344015
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 3720)
AUTHORS Boulaud, A.
TITLE Recombinant laminin 5
JOURNAL Patent: WO 0066731-A 29 09-NOV-2000;
Blastatum, Inc. (US)
FEATURES
source Location/Qualifiers
1..3720
/organism="Homo sapiens"
/db_xref="taxon:9606"
38..3619
/note="unnamed protein product"
/codon_start=1
/protein_id="CAC17367.1"
/db_xref="GI:11344016"
/translation="MPALMLGCCCCFSLILPAARATSRREVCDCKSGKRCITFDRELH
KOTNGFRCLNNDNTDGIHCECKNGFYRHERRCILPCNCSNGSSLSARCDNSGR
SCKPGVTARCDRCILPGRHMLTDAGCTDORLIDSKCPCDPAGISGPDAGRCVCKPA
VTGERCDRCRSGYINLDGNEPGCTQCTCYGHSASCSAETSVKRTISTFDYDGM
KAVONGSPAKLQMSORHODVFSQAORLDPYFVAPAFELGQOVSYQSSPOLSYF
RGRHPSADHVLLEGAGLRTAPLMLDKLTIPCGILTKTFTPLNPNMNSPOLSYF
EYRLRLNLTAIRATYGEYSTGYIDVNTLISAPVSGAPAPVEOCICPVYGGQF
CODCSAGYRKDSARLPGPGTCLPCNCGGACDPPTGDCYSGDENPDIECACPIGFY
NDPHDRCKPCPCCHNGSCSYMPTEBEVYCNCPGVTAGACELCAGYFEDPGEH
GVPKCPCCCNNTDPSASGNDRLTGRCLKICINTAGIYCDQKAGYFEDPGEH
ADKRCACNCPMPSGSEPCVCKRPGEGPCHEGAFSCPCACYNQVKTOMDPM
OOLORMEALISKAGCGDVPTTELEGRMOQAEOLLOPILDAQISEGASRSLQIA
KYRSEENYQSRHLDLKTMTVERVRAIGSOTQORVADTIRLITOMOLSLAESASLGT
NIPADHYVGPNGFKSLAOEATRLAESVESASNEQLTRETEDYSKALSLVRALH
EGVSGSGSPGDAVOGLVERLEKTRSLAQOITRATQAELEADYSKADQSLIDNSG
PLOGVSDSPQVEEAKRIKORADSLSVTRHMDPFTYVETLKNLKEADQSLIDNSG
SGREKSDILSRANLAKSRAOALSGNATFEVSIILKNLEPDILOYDNRKAEAEA
MKRLSISOKYASDADTKTOAERALGSAADQRAKNGCEALFISSELEIGSILN
EANTADALAMEKSLASIKSMREVBELERKELEPTNMDAVQVYTTAEKQVDTRA
KNAGVTTDDTLNLDGLLHMDQPLSVDEBGLVLEQKLSRAKTIINSOLRPMNSELE
ERARQQRHLHLETSIDGILADVKNLENIRDNLPFGCYNTQALEQO"

sig-peptide
BASE COUNT 969 a 900 c 1067 g 784 t
ORIGIN

Query Match 92.0%; Score 18.4; DB 6; Length 3720;
Best Local Similarity 95.0%; Pred. No. 29;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 TGTACTCAGGGATGAGAA 20
|||||
Db 1328 TGTATTCAGGGATGAGAA 1347

RESULT 5

162750
LOCUS 162750 4316 bp DNA linear PAT 07-OCT-1997
DEFINITION Sequence 14 from patent US 5660982.
ACCESSION 162750
VERSION 162750.1 GI:2480458
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 4316)
AUTHORS Trygvason, K., Kallunki, P. and Pyke, C.
TITLE Laminin chains: diagnostic uses
JOURNAL Patent: US 5660982-A 14 26-AUG-1997;
Location/Qualifiers
FEATURES
source 1..4316
/organism="unknown"
BASE COUNT 1158 a 1033 c 1226 g 899 t
ORIGIN

Query Match 92.0%; Score 18.4; DB 6; Length 4316;
Best Local Similarity 95.0%; Pred. No. 28;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 TGTACTCAGGGATGAGAA 20
|||||
Db 1408 TGTATTCAGGGATGAGAA 1427

RESULT 6
LOCUS HSLAMB2TB 4316 bp mRNA linear PRI 27-MAR-1996
DEFINITION H.sapiens mRNA for laminin.
ACCESSION Z15009 S47253
VERSION Z15009.1 GI:34231
KEYWORDS laminin.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 4316)
AUTHORS Trygvason, K.
TITLE Direct Submission
JOURNAL Submitted (27-AUG-1992) Trygvason K., Biocenter and University of
Oulu, Biochemistry, Linnaumaa, Oulu, Finland, SF-90570
2 (bases 1 to 4316)
Kallunki, P., Sahlro, K., Eddy, R., Byers, M., Kallunki, T., Sariola, H.,
Beck, K., Hirvonen, H., Shows, T. B. and Trygvason, K.
A truncated laminin chain homologous to the B2 chain: structure,
spatial expression, and chromosomal assignment
J. Cell Biol. 119 (3), 679-693 (1992)
J. Cell Biol. 93016279
1383240
FEATURES
source Location/Qualifiers
1..4316
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="1"
/cell_type="Fibroblastoma"
/cell_type="Fibroblastoma"
118..3453
/codon_start=1
/product="laminin"
/protein_id="CAA78729.1"
/db_xref="GI:34232"
/db_xref="SWISS-PROT:Q13753"
/translation="MPALMLGCCCCFSLILPAARATSRREVCDCKSGKRCITFDRELH
KOTNGFRCLNNDNTDGIHCECKNGFYRHERRCILPCNCSNGSSLSARCDNSGR
SCKPGVTARCDRCILPGRHMLTDAGCTDORLIDSKCPCDPAGISGPDAGRCVCKPA
VTGERCDRCRSGYINLDGNEPGCTQCTCYGHSASCSAETSVKRTISTFDYDGM
KAVONGSPAKLQMSORHODVFSQAORLDPYFVAPAFELGQOVSYQSSPOLSYF
RGRHPSADHVLLEGAGLRTAPLMLDKLTIPCGILTKTFTPLNPNMNSPOLSYF
EYRLRLNLTAIRATYGEYSTGYIDVNTLISAPVSGAPAPVEOCICPVYGGQF
CODCSAGYRKDSARLPGPGTCLPCNCGGACDPPTGDCYSGDENPDIECACPIGFY

CDS

NDPHRSCPCPCNHNFGSCSVYPTETEEVVCNCPGVTGANCCELADGYPGDPGEH
GVRRCQPCQCNNDVNDPSAGNCDRITGCLCINHTAGIYCDQCKAGYFGLPAPNP
ADKCRACNPNMPSGSEPCGSDGCVCGPCGPGSCPCAGCPCACVNOVKIOMOPM
QOLQRMALISAKOAGDGVPTPELEGRMOQAEOALQDILRAQISEGASRSLQILA
KVRQSENTYQSRHLDLAKMTVERVALGSOYQKRVYDTHLITLITOMLSLAESASIGNT
NIPADHYVGPNGFKSLAOEATRLAESHSNNEQTLREDEYSKALSLVRALH
EYVGSQSPDPAVOVQGLVEKLEKTKSLAOQLTRATOAEIADSYOHLIDNGS
PLOGVSDSPVEAKRIKOKADSLSYTRHMDERKTROKNIGMKMEBAQOOLLNGK
SGREKSDILSRANLAKSRAOALSMGNAPEVESEILKNREPLLOYDNRAKAEBA
MKRLSYISOKVSDADSKTOQARALGSAADQARAKNGAGELTSESLEQDITSLA
EANTVADGALAEKGLASIKSREVESELEKLEFDTNMAVQVMTTEAKVDTTRA
KNAGVTIDTNTLDTGLHLIMGM"
sig_peptide
repeat_unit
4021..4316
/rpt_family="Human ALU"
/rpt_type=OTHER
BASE COUNT 1158 a 1033 c 1226 g 899 t
ORIGIN

Query Match 92.0%; Score 18.4; DB 9; Length 4316;
Best Local Similarity 95.0%; Pred. No. 28;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 TGTACTCAGGGGATGAGAA 20
||||| |||||||
Db 1408 TGTATTCTAGGGGATGAGAA 1427

RESULT 7
LOCUS AX045563 5020 bp DNA linear PAT 24-NOV-2000
DEFINITION Sequence 27 from Patent W00066731.
ACCESSION AX045563
VERSION AX045563.1 GI:11344013
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE 1 (bases 1 to 5020)
AUTHORS Boulaud, A.
TITLE Recombinant laminin 5
JOURNAL Patent: WO 0066731-A 27 09-NOV-2000;
BioStalium, Inc. (US)
FEATURES
source Location/Qualifiers
1..5020
/organism="Homo sapiens"
/db_xref="taxon:9606"
<1..3519
/note="unnamed protein product"
/codon_start=1
/protein_id="CAC17366.1"
/db_xref="GI:11344014"
//translation="MSRRPYCDCKNGKSRQCIIDREHLHRTGNGFCLNCNDMTDITHC
EKCRNGFTYRHRDRKCLPCNCKSKSLKRCNDSKRCPEVTARCDKCLPGHML
TDAGCTQDQRLDSKCDPAGIAGPCAGRCVCKPAVTEGDCRSQYLMQDGNP
EECTQCFYGHSAKSCSSAAYSVHRTSTFHQDVGMVAVNGSPAKIOMQORDY
FSSAQRLDPYVAPAKFELGNOQVSYGQSLSDYVDRGRHPSADHYVLEAGRTIT
APIMPLGTLIPGLTKTYTERLNEHPSNNMSQQLSYFEYRLRLNLTLLRATAGEY
STGYIDNTLISARVSGAPAPVBOCICPVGKQFOCDQCSGYKRSARLPGPGCT
IPCNCQGGACDPDPTGDCYSGDENPDIEADCPITFTRPHDPHRCCKPCPCNGSCS
VLPETEYVVCNCPGVTGARCELCADGTFGDFEHGVRPCQPCQCNNDVPSAGS
NCDRLTGKCLKCIHNTAGIYCDQCKAGYFGLPAPNPADKCRACNPNMPSGSCRS
DGTVCCKPGFGPCNCEHAGFSCPCACYNQVKIOMDPMOOLQMEALISAKOAGDGVPT
PELEGRMOQAEOALQDILRAQISEGASRSLQILAESHSNNEQTLREDEYSKALSLVRALH
EYVGSQSPDPAVOVQGLVEKLEKTKSLAOQLTRATOAEIADSYOHLIDNGS
PLOGVSDSPVEAKRIKOKADSLSYTRHMDERKTROKNIGMKMEBAQOOLLNGK
SGREKSDILSRANLAKSRAOALSMGNAPEVESEILKNREPLLOYDNRAKAEBA
MKRLSYISOKVSDADSKTOQARALGSAADQARAKNGAGELTSESLEQDITSLA
EANTVADGALAEKGLASIKSREVESELEKLEFDTNMAVQVMTTEAKVDTTRA
KNAGVTIDTNTLDTGLHLIMGM"

CDS
1..5020
/organism="Homo sapiens"
/db_xref="taxon:9606"
<1..3519
/note="unnamed protein product"
/codon_start=1
/protein_id="CAC17366.1"
/db_xref="GI:11344014"
//translation="MSRRPYCDCKNGKSRQCIIDREHLHRTGNGFCLNCNDMTDITHC
EKCRNGFTYRHRDRKCLPCNCKSKSLKRCNDSKRCPEVTARCDKCLPGHML
TDAGCTQDQRLDSKCDPAGIAGPCAGRCVCKPAVTEGDCRSQYLMQDGNP
EECTQCFYGHSAKSCSSAAYSVHRTSTFHQDVGMVAVNGSPAKIOMQORDY
FSSAQRLDPYVAPAKFELGNOQVSYGQSLSDYVDRGRHPSADHYVLEAGRTIT
APIMPLGTLIPGLTKTYTERLNEHPSNNMSQQLSYFEYRLRLNLTLLRATAGEY
STGYIDNTLISARVSGAPAPVBOCICPVGKQFOCDQCSGYKRSARLPGPGCT
IPCNCQGGACDPDPTGDCYSGDENPDIEADCPITFTRPHDPHRCCKPCPCNGSCS
VLPETEYVVCNCPGVTGARCELCADGTFGDFEHGVRPCQPCQCNNDVPSAGS
NCDRLTGKCLKCIHNTAGIYCDQCKAGYFGLPAPNPADKCRACNPNMPSGSCRS
DGTVCCKPGFGPCNCEHAGFSCPCACYNQVKIOMDPMOOLQMEALISAKOAGDGVPT
PELEGRMOQAEOALQDILRAQISEGASRSLQILAESHSNNEQTLREDEYSKALSLVRALH
EYVGSQSPDPAVOVQGLVEKLEKTKSLAOQLTRATOAEIADSYOHLIDNGS
PLOGVSDSPVEAKRIKOKADSLSYTRHMDERKTROKNIGMKMEBAQOOLLNGK
SGREKSDILSRANLAKSRAOALSMGNAPEVESEILKNREPLLOYDNRAKAEBA
MKRLSYISOKVSDADSKTOQARALGSAADQARAKNGAGELTSESLEQDITSLA
EANTVADGALAEKGLASIKSREVESELEKLEFDTNMAVQVMTTEAKVDTTRA
KNAGVTIDTNTLDTGLHLIMGM"

DOPLSVDEGLVTLLEOKLSRANKQINSOLRPMMSLEBERARQOCHLLETSIDGIL
ADVKNLENTRDMLPGCYNTOLAEQO"
BASE COUNT 1329 a 1170 c 1333 g 1188 t
ORIGIN

Query Match 92.0%; Score 18.4; DB 6; Length 5020;
Best Local Similarity 95.0%; Pred. No. 28;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 TGTACTCAGGGGATGAGAA 20
||||| |||||||
Db 1228 TGTATTCTAGGGGATGAGAA 1247

RESULT 8
LOCUS AX365737 5156 bp DNA linear PAT 15-FEB-2002
DEFINITION Sequence 130 from Patent W00200174.
ACCESSION AX365737
VERSION AX365737.1 GI:18697286
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

REFERENCE 1
AUTHORS Wang, T., Wang, A., Skeiky, Y.A., Li, S.X., Kalos, M.D., Henderson, R.A.,
McNeill, P.D., Fanger, N., Retter, M.W., Marnerakis, M., Fanger, G.R.,
Vedlick, T.S., Carter, D., Wattanabe, Y., and Peckham, D.W.
TITLE Compositions and methods for the therapy and diagnosis of lung
cancer
JOURNAL Patent: WO 0200174-A 130 03-JAN-2002;
CORIXA CORPORATION (US)
FEATURES
source Location/Qualifiers
1..5156
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 1351 a 1222 c 1377 g 1206 t
ORIGIN

Query Match 92.0%; Score 18.4; DB 6; Length 5156;
Best Local Similarity 95.0%; Pred. No. 28;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 TGTACTCAGGGGATGAGAA 20
||||| |||||||
Db 1356 TGTATTCTAGGGGATGAGAA 1375

RESULT 9
LOCUS HSNICE 5156 bp mRNA linear PRI 02-FEB-1994
DEFINITION H.sapiens mRNA for niclein B2 chain.
ACCESSION X73902
VERSION X73902.1 GI:452754
KEYWORDS Basement membrane; laminin; niclein; structural protein.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE 1 (bases 1 to 5156)
AUTHORS Vailly, J., Verrando, P., Champilaud, M.F., Gerecke, D., Wagman, D.W.,
Baudoin, C., Aberdam, D., Burgesson, R., Bauer, E. and Ortonne, J.P.
TITLE The 100-kDa chain of niclein/kallinin is a laminin B2 chain variant
JOURNAL Eur. J. Biochem. 219 (1-2), 209-218 (1994)
MEDLINE 94139694
PUBMED 8306988
REFERENCE 2 (bases 1 to 5156)
AUTHORS Vailly, J.
TITLE Direct Submission
JOURNAL Submitted (01-JUL-1993) J. Vailly, Lab de Recherches
Dermatologiques, Av De Valombrese Faculte de Medecine, 06107 Nice,
Cedex 2, FRANCE

COMMENT Related sequence: Z15008.
FEATURES Location/Qualifiers
SOURCE 1..5156
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="1"
/map="1q25-31; in situ hybrid"
/cell_type="keratinocytes"
66..3647
/note="B2 chain; 100 kDa subunit"
/codon_start=1
/product="nicotin"
/protein_id="CA52108.1"
/db_xref="GI:452755"
/db_xref="SWISS-PROT:Q13753"
/translation="MPALMLGCCCLCSLLPAAATSRREYCDGCKSGROCIIPRELI
RQNGNFRCLNCNDNTDGIHECKNGFYRHRERDCLPCNCKNSGSLACDNGSGC
SCKPGYVGARCDRLPGFHMFTDAGCTODDLSDCKDCDPAGIAGPDAACVCKPA
VTGERCDRCRGYNIIDGNGPECTOCFCYSHASCSASAEYVHKITSTHODVGA
KAVORNGSPAKLQWSORHODVSSAORLDPVYFAPAKFLGNOQVSYGQSLSPFYRD
RGRHPSAHVILLEGAGLRTAPLPIGLKTLPCGLITTYFRLMEHPSNMNSPOLSYF
EYRLRLNLRLALRIATYGERSTGYINVTILISAPVSGAPAPVVEQICVYKGGF
CODASGKRDASRLPGPTCTPCNCGGACDPTDGYSGDENPDIECADCPIGFY
NDPDPSCRCPCPNHGFSCSVMPETEVEVCNCPGVTGARCELCADGYDEPGEH
GVPAPCOPCQNNNVDSASGNCRLTGRCIKCIHNTAGIYCDCKAGYFEDPLAPN
ADKRCACNCPMSEPEVGCSDGTVCCKPGFGCEHAGSPACYNQVITQMDQEM
QOOLORMEALISKAGGQGVVDTELEGMOQAEALDILDAQISGASASLQIQA
KYRSQENSYOSRLDILKMTYRVRALISOYQNRVDRHLITQMLSLAESASLGNT
NIPASDHVGPNGFSLAQLAEATRLAESVESASMMEDLTRETEYSKQALSLVKALH
EGVSGSGSPGAVOGLVEKLEKTLAEQSLAQLTREAQAEALTEADRSYQSLRLDSYS
RLQGVSDQSEFVEEAKRIKQADLSLVTREHMEFRKQNLNWKMEEAQQLQNGK
SGRKSQDLISRAVLAKSRAQEAALSMGNATFEYESILKLNREEDLOYDNKAAEEA
MKRLSYISQKVSADSKTQOAEALISMGATFEYESILKLNREEDLOYDNKAAEEA
EANTYADALAMEGLASLSKSEMEVEBELKELEPTNDAYQVITTEAKQVDTA
KNAQVTTQDITNTLDGILLHMDPLSYDEBELVLEQLSRAKQIINSQLRPMSELE
ERRAQQRGHLHLETSIDGILLADYKLENLINDNLPQCYNTQALEQO"

BASE COUNT 1351 a 1222 c 1377 g 1206 t
ORIGIN

Query Match 92.0%; Score 18.4; DB 9; Length 5156;
Best Local Similarity 95.0%; Pred. No. 28;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TGTACTCAGGGGATGAGAA 20
|||||
Db 1356 TGTATTTCAGGGGATGAGAA 1375

RESULT 10
AX045561 5200 bp DNA linear PAT 24-NOV-2000
LOCUS DEFINITION Sequence 25 from Patent W00066731.
ACCESSION AX045561
VERSION AX045561.1 GI:11344011
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 5200)
AUTHORS Boutaud,A.
TITLE Recombinant laminin 5
JOURNAL Patent: WO 0066731-A 25 09-NOV-2000;
BioStatum, Inc. (US)
FEATURES Location/Qualifiers
SOURCE 1..5200
/organism="Homo sapiens"
/db_xref="taxon:9606"
118..3699
/note="unnamed protein product"
/codon_start=1
/protein_id="CAC17365.1"
/db_xref="GI:11344012"

/translation="MPALMLGCCCLCSLLPAAATSRREYCDGCKSGROCIIPRELI
RQNGNFRCLNCNDNTDGIHECKNGFYRHRERDCLPCNCKNSGSLACDNGSGC
SCKPGYVGARCDRLPGFHMFTDAGCTODDLSDCKDCDPAGIAGPDAACVCKPA
VTGERCDRCRGYNIIDGNGPECTOCFCYSHASCSASAEYVHKITSTHODVGA
KAVORNGSPAKLQWSORHODVSSAORLDPVYFAPAKFLGNOQVSYGQSLSPFYRD
RGRHPSAHVILLEGAGLRTAPLPIGLKTLPCGLITTYFRLMEHPSNMNSPOLSYF
EYRLRLNLRLALRIATYGERSTGYINVTILISAPVSGAPAPVVEQICVYKGGF
CODASGKRDASRLPGPTCTPCNCGGACDPTDGYSGDENPDIECADCPIGFY
NDPDPSCRCPCPNHGFSCSVMPETEVEVCNCPGVTGARCELCADGYDEPGEH
GVPAPCOPCQNNNVDSASGNCRLTGRCIKCIHNTAGIYCDCKAGYFEDPLAPN
ADKRCACNCPMSEPEVGCSDGTVCCKPGFGCEHAGSPACYNQVITQMDQEM
QOOLORMEALISKAGGQGVVDTELEGMOQAEALDILDAQISGASASLQIQA
KYRSQENSYOSRLDILKMTYRVRALISOYQNRVDRHLITQMLSLAESASLGNT
NIPASDHVGPNGFSLAQLAEATRLAESVESASMMEDLTRETEYSKQALSLVKALH
EGVSGSGSPGAVOGLVEKLEKTLAEQSLAQLTREAQAEALTEADRSYQSLRLDSYS
PLQGVSDQSEFVEEAKRIKQADLSLVTREHMEFRKQNLNWKMEEAQQLQNGK
SGRKSQDLISRAVLAKSRAQEAALSMGNATFEYESILKLNREEDLOYDNKAAEEA
MKRLSYISQKVSADSKTQOAEALISMGATFEYESILKLNREEDLOYDNKAAEEA
EANTYADALAMEGLASLSKSEMEVEBELKELEPTNDAYQVITTEAKQVDTA
KNAQVTTQDITNTLDGILLHMDPLSYDEBELVLEQLSRAKQIINSQLRPMSELE
ERRAQQRGHLHLETSIDGILLADYKLENLINDNLPQCYNTQALEQO"

BASE COUNT 1364 a 1236 c 1392 g 1208 t
ORIGIN

sig.peptide
BASE COUNT 1364 a 1236 c 1392 g 1208 t

Query Match 92.0%; Score 18.4; DB 6; Length 5200;
Best Local Similarity 95.0%; Pred. No. 28;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TGTACTCAGGGGATGAGAA 20
|||||
Db 1408 TGTATTTCAGGGGATGAGAA 1427

RESULT 11
162749 5200 bp DNA linear PAT 07-OCT-1997
LOCUS DEFINITION Sequence 12 from patent US 5660982.
ACCESSION 162749
VERSION 162749.1 GI:2480457
KEYWORDS Unknown.
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 5200)
AUTHORS Tryggyvason,K., Kallunki,P. and Pyke,C.
TITLE Laminin chains: diagnostic uses
JOURNAL Patent: US 5660982-A 12 26-AUG-1997;
FEATURES Location/Qualifiers
SOURCE 1..5200
/organism="unknown"
BASE COUNT 1364 a 1236 c 1392 g 1208 t
ORIGIN

Query Match 92.0%; Score 18.4; DB 6; Length 5200;
Best Local Similarity 95.0%; Pred. No. 28;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TGTACTCAGGGGATGAGAA 20
|||||
Db 1408 TGTATTTCAGGGGATGAGAA 1427

RESULT 12
HSLAMB2T 5200 bp mRNA linear PRI 27-MAR-1996
LOCUS DEFINITION H. sapiens mRNA for laminin.
ACCESSION Z15008 S47028
VERSION Z15008.1 GI:34229
KEYWORDS laminin.
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE 1 (bases 1 to 5200)
 Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 Kallunki, P., Sainio, K., Eddy, R., Byers, M., Kallunki, T., Sariola, H., Beck, K., Hirvonen, H., Shows, T. B., and Tryggvason, K.
 A truncated laminin chain homologous to the B2 chain: structure, spatial expression, and chromosomal assignment
 J. Cell Biol. 119 (3), 679-693 (1992)

JOURNAL MEDLINE
 PUBMED 1383240
 2 (bases 1 to 5200)
 Tryggvason, K.
 Direct Submission
 Submitted (27-AUG-1992) Tryggvason K., Biocenter and University of Oulu, Biochemistry, Linnanmaa, Oulu, Finland, SF-90570
 Location/Qualifiers

FEATURES
 source
 1. 5200
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="1"
 /cell_line="Ht-1080"
 /cell_type="Fibroblastoma"
 118. 3699
 /codon_start=1
 /product="Laminin"
 /protein_id="CAA78728.1"
 /db_xref="GI:34230"
 /db_xref="SWISS-PROT:O13753"
 /translation="MPALMIGCCICFSLLPAPARATRRREVCDCNGKSRQCFDEBELH
 ROTNGFRCLNNDNDGICHCCKNGFRHREDRCLCNCNSNGSDPCADGNCGRK
 SCPCVTGARDCRCLPGPHMLDAGCTODRLLDCKDCDDPCDPCADGNCGRK
 VTGERCDRERSYVNLGDGNGPCGTCQCFYGHASACRSASRYHKTSTFQDDYD
 KAVQNGSPAKLQMSQRHODVSSAQRDLPVFAVPAKLGQVQVSYGSLSTFQDDYD
 RGRHPSADHVLLEAGALITAPLMPLGCTGKTKYTFPLNHEHSPNSPOLSYF
 EKRRLRLNTALIRATYGEVSTGYIDNLTLSARVSGAPAPWEOCIPVYGOF
 CDDCASGYKRDARLPGFCTCIPCNCGGACDPPGDCSGDENPDIACDQPIGTF
 NDHPDRSCKPCGCHNGFSCSVIPPEEVEVYNNCPGVTGAGCAICADGREDPGEH
 GVRPCRCQCCNSNTPRPSAGNCDRLTGCLCKTINTAGITDCKACATGCPPLANP
 ADKCRACNCPNGSEPRGCRSDGTCVCKRFGEGPNCHEGAFSCPCYNOKTOMQFM
 OOLQRMELISKAQGDGVPTLEGRMQQAOLADLINAQISEGASRLGIQFL
 KVRQENSYOSRLDKMTVERVALGSOYQNRVDTHTLITLOMQLSLAESASIGNT
 NTPADHYGPNFGSLQAEATRLAESHYESASNMQLREDEYSKQSLSVRRALH
 EGVSGSGPDGAVVOGLYEKLEKTRLSAQOLTRFATOETIADRSYOHSLNVS
 PEGVSDSGFQVEAFKRIKORADSLSLVTRHMDKFRKOKLWKKMEKRAOOLNKG
 SRRKSDLSLRANAKSKAQAALSKGNATFEVESILKNEFLDLYDNKRAEAE
 MKRLSYISKVSDASDKTQOARALGSAADQRAKNGAGEALISSELEQIGSLN
 EKNVADGALAMEKGLASLSKSEMEVEGELERLEFDNMDAVOMVITEAKVYTRA
 KNAGVYIQTTLTLDGLHMDQPLSVDEEGIVLEOKLSRAKTOINSQLRPMSELE
 ERARQOGLHLETSIDGILADVKNLENIRNLPDPCYNTQALEQO"

sig_peptide
 4433
 5195
 polyA_site
 BASE COUNT 1364 a 1236 c 1392 g 1208 t
 ORIGIN

Query Match 92.0%; Score 18.4; DB 9; Length 5200;
 Best Local Similarity 95.0%; Pred. No. 28;
 Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 TGTACTCAGGAGATGAGAA 20
 ||||| ||||| ||||| |||||
 Db 1408 TGTATTTCAGGAGATGAGAA 1427

RESULT 13
 AC131536
 LOCUS AC131536/c
 DEFINITION Rattus norvegicus clone CH230-49E2, *** SEQUENCING IN PROGRESS ***,
 48 unordered pieces.
 AC131536 98233 bp DNA linear HTG 24-AUG-2002
 VERSION AC131536.1 GI:22474800
 KEYWORDS HTG; HTGS_PHASE1.
 SOURCE Norway rat.
 ORGANISM Rattus norvegicus

REFERENCE 1 (bases 1 to 98233)
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.
 Muzny, D. Marie, Metzker, M. Lee, Abramson, S., Adams, C., Alder, J.,
 Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D.,
 Anyalebech, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H.,
 Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benhammed, F.,
 Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,
 Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E.,
 Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A.,
 Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J.,
 Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L.,
 Davila, M. L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D.,
 Delgado, O., Denison, S., Deramo, C., Ding, Y., Din, H., Divya, K.,
 Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K.,
 Egan, A., Escotto, M., Eugene, C., Evans, C. A., Falls, T., Fan, G.,
 Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,
 Fraser, C. M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M.,
 Gebregeorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W.,
 Gunaratne, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, K.,
 Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J.,
 Hernandez, R., Hines, S., Hladun, S. L., Hodgson, A., Hognes, M.,
 Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A.,
 Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolyet, A.,
 Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C.,
 Kowis, C., Kraft, C. L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, D.,
 Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J.,
 Lorusubhe, L., Louisedge, H., Lozdo, R. J., Lu, X., Ma, J.,
 Maheshwari, M., Mahindaratne, M., Mahmoud, M., Malloy, K., Mangum, A.,
 Mangum, B., Mapa, P., Martin, K., Martin, R., Martinez, E.,
 Mawlin, S., McLeod, M., McNeill, T., Meenan, E., Milosavljevic, A.,
 Miner, G., Ming, E., Montemayor, J., Moore, S., Morgan, M., Morris, K.,
 Morris, S., Mundada, M., Murphy, M., Natr, L., Nankervis, C., Neal, D.,
 Newton, N., Nguyen, N., Norris, S., Nwackelme, O., Okunou, G.,
 Olarunpungson, A., Pal, S., Parks, K., Pasternak, S., Paul, H.,
 Perez, A., Perez, L., Pfannkuch, C., Plopper, F., Poindester, A.,
 Popovic, D., Primus, E., Pu, L., L., Puazo, M., Quintoz, J., Raculin, E.,
 Reeves, K., Regier, M. A., Reigh, R., Reilly, B., Reilly, M., Ren, Y.,
 Reuter, M., Richards, S., Riggs, F., Rivers, C., Rodkey, T., Rojas, A.,
 Rose, M., Rose, R., Ruiz, S. J., Sanders, W., Savery, G., Scherer, S.,
 Scott, G., Shatsman, S., Shen, H., Shetty, J., Shvartsbeyn, A.,
 Slason, I., Sitter, C. D., Smaj, D., Sneed, A., Sodergren, E.,
 Song, X., Z., Sorelle, R., Sosa, J., Steinhil, M., Strong, R., Sutton, A.,
 Svatek, A., Tabori, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S.,
 Tingey, A., Tiejos, Z., Usmani, K., Valas, R., Vera, V., Villasana, D.,
 Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J.,
 Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wleczky, R.,
 Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S.,
 Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X.,
 Zhao, S., Zhou, D., von Niederhausen, A., Weiss, R., Smith, D. R.,
 Holt, R. A., Smith, H. O., Weinstock, G., and Gibbs, R. A.
 Direct Submission
 Unpublished
 2 (bases 1 to 98233)
 Rat Genome Sequencing Consortium.
 Direct Submission
 Submitted (24-AUG-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 ----- Genome Center
 Center: Baylor College of Medicine
 Center code: BCM
 Web site: http://www.hgsc.bcm.tmc.edu/
 Contact: hgsc-help@bcm.tmc.edu
 ----- Project Information
 Center project name: GRUR
 Center clone name: CH230-49E2
 ----- Summary Statistics
 Sequencing vector: Plasmid;
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.990329
 Consensus quality: 47238 bases at least Q40

Consensus quality: 52180 bases at least Q30
Consensus quality: 55537 bases at least Q20

NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
NOTE: This is a 'working draft' sequence. It currently
consists of 48 contigs. The true order of the pieces
is not known and their order in this sequence record is
arbitrary. Gaps between the contigs are represented as
runs of N, but the exact sizes of the gaps are unknown.
This record will be updated with the finished sequence
as soon as it is available and the accession number will
be preserved.

1 1046: contig of 1046 bp in length
1047 1146: gap of unknown length
1147 2583: contig of 1437 bp in length
2584 2683: gap of unknown length
2684 3721: contig of 1038 bp in length
3722 3821: gap of unknown length
3822 5429: contig of 1608 bp in length
5430 5529: gap of unknown length
5530 6756: contig of 1227 bp in length
6757 6856: gap of unknown length
6857 7887: contig of 1031 bp in length
7888 9231: contig of 1244 bp in length
9232 9331: gap of unknown length
9332 10413: contig of 1082 bp in length
10414 10513: gap of unknown length
10514 12049: contig of 1536 bp in length
12050 12149: gap of unknown length
12150 13250: contig of 1101 bp in length
13251 13350: gap of unknown length
13351 14619: contig of 1269 bp in length
14620 14719: gap of unknown length
14720 15729: contig of 1010 bp in length
15730 15829: gap of unknown length
15830 16966: contig of 1137 bp in length
16967 17066: gap of unknown length
17067 18149: contig of 1083 bp in length
18150 18249: gap of unknown length
18250 19642: contig of 1393 bp in length
19643 19742: gap of unknown length
19743 21226: contig of 1484 bp in length
21227 22596: gap of unknown length
22597 22696: contig of 1270 bp in length
22697 23759: contig of 1063 bp in length
23760 23859: gap of unknown length
23860 25371: contig of 1512 bp in length
25372 25471: gap of unknown length
25472 26743: contig of 1272 bp in length
26744 26843: gap of unknown length
26844 28253: contig of 1410 bp in length
28254 30556: gap of unknown length
30557 30656: contig of 2203 bp in length
30657 32460: gap of unknown length
32461 32560: contig of 1804 bp in length
32561 33609: gap of unknown length
33610 33709: contig of 1049 bp in length
33710 35142: contig of 1433 bp in length
35143 35242: gap of unknown length
35243 36850: contig of 1608 bp in length
36851 36950: gap of unknown length
36951 38992: contig of 1942 bp in length
38993 38992: gap of unknown length
38993 40317: contig of 1325 bp in length
40318 40417: gap of unknown length
40419 41622: contig of 1205 bp in length
41623 41722: gap of unknown length
41723 43934: contig of 2212 bp in length
43935 44034: gap of unknown length

44035 45686: contig of 1652 bp in length
45687 45786: gap of unknown length
45787 47335: contig of 1549 bp in length
47336 47435: gap of unknown length
47436 49695: contig of 2260 bp in length
49696 49795: gap of unknown length
49796 52881: contig of 3086 bp in length
52882 52981: gap of unknown length
52982 55039: contig of 2058 bp in length
55040 55139: gap of unknown length
55140 57295: contig of 2156 bp in length
57296 57395: gap of unknown length
57396 59052: contig of 1657 bp in length
59053 59152: gap of unknown length
59153 61795: contig of 2643 bp in length
61796 61895: gap of unknown length
61896 64192: contig of 2297 bp in length
64193 64292: gap of unknown length
64293 66929: contig of 2637 bp in length
66930 67029: gap of unknown length
67030 69943: contig of 2914 bp in length
69944 70043: gap of unknown length
70044 72183: contig of 2140 bp in length
72184 72283: gap of unknown length
72284 75162: contig of 2879 bp in length
75163 75262: gap of unknown length
75263 79581: contig of 4319 bp in length
79582 79681: gap of unknown length
79682 82554: contig of 2873 bp in length
82555 82654: gap of unknown length
82655 86943: contig of 4289 bp in length
86944 87043: gap of unknown length
87044 93641: contig of 6598 bp in length
93642 93741: gap of unknown length
93742 98233: contig of 4492 bp in length.

FEATURES
source 1..98233
/organism="Rattus norvegicus"
/db_xref="taxon:10116"
/clone="CH230-49E2"

BASE COUNT 24506 a 21274 c 21399 g 24747 t 6307 others
ORIGIN

Query Match 92.0%; Score 18.4; DB 2; Length 98233;
Best Local Similarity 95.0%; Pred. No. 20;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 TGTTCAGGAGGATGAGAA 20
|||||
DB 30307 TGTTCAGGAGGATGAGAA 30288

RESULT 14
HSDJ383J4 133936 bp DNA linear PRI 30-SEP-2000
LOCUS HSDJ383J4
DEFINITION Human DNA sequence from clone Rp3-383J4 on chromosome 1q24.1-24.3
Contains part of a gene encoding a kelch motif containing protein,
part of a novel gene encoding a protein similar to Aspartyl-tRNA
synthetase, a putative novel gene, a 40S ribosomal protein 527
(RPS27) pseudogene, 2 Cpg islands, ESTs, STSS and GSSs, complete
sequence.

ACCESSION AL109921.21 GI:9586678
KEYWORDS HTG: Aspartyl-tRNA synthetase; Cpg island; kelch-motif; RPS27.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 133936)
AUTHORS Frankland, J.
TITLE Direct Submisslon
JOURNAL Submitted (25-SEP-2000) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone

COMMENT

requests: clonerequest@sanger.ac.uk
On Jul 31, 2000 this sequence version replaced 91:9581776.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence has been finished according to sequence map criteria
as follows. An attempt is made to resolve all sequencing problems,
such as compressions and repeats, but not necessarily within known
annotated human repeat sequence elements (e.g. Alu). Where the
sequence is ambiguous, there is an annotation using the 'unsure'
feature key.

FEATURES

Source

```

repeat_region 339..6438
/note="L1BP4 repeat: matches 5..6142 of consensus"
6490..6784
/note="AluY repeat: matches 1..295 of consensus"
repeat_region 6788..6916
/note="FLAM A repeat: matches 6..142 of consensus"
misc_feature complement(7412..7836)
/note="match: SMS: Em:HSP16f10"
7396..7888
/note="AluXc repeat: matches 1..302 of consensus"
repeat_region 8040..8172
/note="FLAM C repeat: matches 1..133 of consensus"
gene join(8287..8860,26337..26495,27786..27880,30496..30611,
31549..31732,40719..40862,48878..49011,50207..50415,
56696..56802,59735..60078)
/note="dJ3833j4.1"
mRNA join(8287..8860,26337..26495,27786..27880,30496..30611,
31549..31732,40719..40862,48878..49011,50207..50415,
56696..56802,59735..60078)
/note="dJ3833j4.1"
CDS /product="dJ3833j4.1 (A Kelch motif-containing protein)"
/note="match: CDNAs: Em:AB026190 Em:AK001430"
/evidence="not experimental"
join(<8287..8860,26337..26495,27786..27880,30496..30611,
31549..31732,40719..40862,48878..49011,50207..50415,
56696..56802,59735..59819)
/note="dJ3833j4.1"
/note="match: proteins: Tr:Q9YVU5 Tr:Q9Y2M5 Tr:O61795
Tr:Q9Y6V6"
/codon_start=2
/evidence="not experimental"
/product="dJ3833j4.1 (A Kelch motif-containing protein)"
/protein_id="CAC10469.1"
/db_xref="GI:12314036"
/translation="CTINIRGEGKMDYTSRCITGDPMILPSGVQPARMPYISDKHR
OTLEVINLKRHELCDDVALVVGAKITTAHTYISAGSPYRPMFTGLASRTQEV
IRIDIDEAMELLIDFYTSQIVVEEGNVOTLLPACILQIETIDACECEFLKRDLS

```

[illegible]


```

repeat_region /note="match: STS: Em:G05508"
16794..16949 /note="LMC/D repeat: matches 5497..5656 of consensus"
repeat_region 17170..17257 /note="44 copies 2 mer at 73% conserved"
repeat_region 17260..17359 /note="50 copies 2 mer ta 75% conserved"
repeat_region 17475..17790 /note="LMC/D repeat: matches 5127..5403 of consensus"
repeat_region 17826..18023 /note="LMC4 repeat: matches 7773..7975 of consensus"
misc_feature /note="match: GSS: Em:A0621823"
18377..19732 /note="LMND1 repeat: matches 4372..5867 of consensus"
repeat_region 19733..20035 /note="Alus repeat: matches 1..312 of consensus"
repeat_region 20036..20446 /note="LMND1 repeat: matches 3983..4372 of consensus"
repeat_region 20449..20547 /note="Alusg repeat: matches 193..291 of consensus"
repeat_region 20564..20972 /note="LMND1 repeat: matches 3522..3974 of consensus"
repeat_region 21028..21531 /note="LMND1 repeat: matches 2914..3451 of consensus"
repeat_region 21851..22253 /note="L1 repeat: matches 2110..2515 of consensus"
repeat_region 22393..23023 /note="L1M4 repeat: matches 3134..3809 of consensus"
repeat_region 23235..23321 /note="L1P4 repeat: matches 6056..6142 of consensus"
repeat_region 23359..23459 /note="L1M2 repeat: matches 6021..6122 of consensus"
repeat_region 24526..24602 /note="MER3 repeat: matches 108..188 of consensus"

```

```

Query Match 92.0%; Score 18.4; DB 9; Length 133936;
Best Local Similarity 95.0%; Pred. No. 19;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

OY 1 TGTACTCAGGAGATGAGAA 20
DB 43829 TGTACTCAGGAGATGAGAA 43848

```

```

RESULT 15
AL354953 164084 bp DNA linear PRI 16-NOV-2001
LOCUS Human DNA sequence from clone RP11-181K3 on chromosome 1, complete
DEFINITION sequence.
ACCESSION AL354953
VERSION AL354953.16 GI:16973821
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 164084)
AUTHORS Tracey, A.
TITLE Direct Submission
JOURNAL Submitted (16-NOV-2001) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquerry@sanger.ac.uk; clone requests: clonerequest@sanger.ac.uk
On Nov 17, 2001 this sequence version replaced gi:16304444.
During sequence assembly data is compared from overlapping clones.
While differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=

```

30): an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em, EMBL; SW, SWISSPROT; Tr, TREMBL; Wp, WORMPEP; Information on the WORMPEP database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/Chr1> RP11-181K3 is from the library RPCI-11.1 constructed by the group of Pieper de Jong. For further details see <http://www.choil.org/bacpac/home.htm>

```

FEATURES
Source 1..164084
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="1"
/clone="RP11-181K3"
/clone_11b="RPCI-11.1"
55965
misc_feature /note="Random repeat. Forced join. Gap size estimated to
be approximately 80bp by restriction digest data."
BASE COUNT 46832 a 33365 c 35491 g 48396 t
ORIGIN

```

```

Query Match 92.0%; Score 18.4; DB 9; Length 164084;
Best Local Similarity 95.0%; Pred. No. 19;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

OY 1 TGTACTCAGGAGATGAGAA 20
DB 133542 TGTACTCAGGAGATGAGAA 133561

```

```

Search completed: July 1, 2003, 07:22:57
Job time : 592.641 secs

```


GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 1, 2003, 03:06:29 ; Search time 110.769 seconds

(without alignments)
406.611 Million cell updates/sec

Title: US-10-053-662A-29

Perfect score: 20

Sequence: 1 tftactcagggagatgagaa 20

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_101002:*

1: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:*

2: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:*

3: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:*

4: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:*

5: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:*

6: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:*

7: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:*

8: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:*

9: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:*

10: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:*

11: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:*

12: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:*

13: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:*

14: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:*

15: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:*

16: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:*

17: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:*

18: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:*

19: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:*

20: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:*

21: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:*

22: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:*

23: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:*

24: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	18.4	92.0	580	AAH09229	Human CDNA clone (
2	18.4	92.0	2310	AAH15301	Human CDNA sequenc
3	18.4	92.0	3582	AAO31145	Human laminin 5 CD
4	18.4	92.0	3620	AAO31145	Human laminin 5 CD
5	18.4	92.0	3720	AAO31145	Human laminin 5 CD
6	18.4	92.0	4316	AAH13324	Kallidin/laminin 5
7	18.4	92.0	4316	AAH13324	Laminin gamma-2 ch
8	18.4	92.0	5020	AAO31145	Human laminin 5 CD
9	18.4	92.0	5152	AAZ24627	Human lung tumor a

10	18.4	92.0	5156	21	AAO5866	Human lung cancer-
11	18.4	92.0	5156	24	ABL49085	Human lung tumour
12	18.4	92.0	5200	17	AAH13323	Kallidin/laminin 5
13	18.4	92.0	5200	21	AAO31145	Human laminin 5 CD
14	18.4	92.0	5200	24	AAH12910	Laminin gamma-2 ch
15	16.4	82.0	20066	24	ABL52278	Human chemokine (C
16	16.4	80.0	14748	24	ABK51275	Human chemokine-2, C
17	16.4	80.0	17026	22	AAK70372	Human immune/haema
18	16.4	80.0	22400	22	AAO5885	Human cardiovascular
19	15.8	79.0	793	20	AAH3981	Gastric cancer ass
20	15.8	79.0	1358	23	AAO50883	DNA encoding novel
21	15.8	79.0	1745	22	AAK91274	Human digestive sy
22	15.8	79.0	2392	24	AAH19443	Human CDNA encodin
23	15.8	79.0	12261	22	AAH13653	Genomic DNA sequen
24	15.8	79.0	56632	22	AAK5581	Human immune/haema
25	15.4	77.0	351	22	AAK57605	Human immune/haema
26	15.4	77.0	1317	20	AAZ20164	Bovine pregnancy a
27	15.4	77.0	7591	22	AAH32887	Human genomic DNA
28	15.4	77.0	14540	22	AAK72853	Human immune/haema
29	15.4	77.0	17596	22	AAK72852	Human immune/haema
30	15.4	77.0	23307	23	ABL19466	Drosophila melanog
31	15.4	77.0	31193	23	ABL29416	Drosophila melanog
32	15.2	76.0	47	21	AAZ66815	Human map-related
33	15.2	76.0	303	22	AAH34949	Human musculoskele
34	15.2	76.0	425	22	AAH36224	Human musculoskele
35	15.2	76.0	454	19	AAV60966	Human immune/haema
36	15.2	76.0	499	22	AAK56847	Human immune/haema
37	15.2	76.0	527	21	AAO33004	Arabidopsis thalia
38	15.2	76.0	531	22	AAH192349	Human polynucleoti
39	15.2	76.0	532	22	AAK61141	Human immune/haema
40	15.2	76.0	565	23	AAH56269	DNA encoding novel
41	15.2	76.0	621	23	ABV55264	Human prostate exp
42	15.2	76.0	696	19	AAV60967	Genomic clone Cks1
43	15.2	76.0	700	24	ABO5862	Arabidopsis thalia
44	15.2	76.0	814	23	AAO50868	DNA encoding novel
45	15.2	76.0	864	23	AAO50865	DNA encoding novel

ALIGNMENTS

RESULT 1	AAH09229/c	AAH09229 standard; CDNA; 580 BP.
ID	AAH09229;	
XX	AAH09229;	
AC	AAH09229;	
XX	AAH09229;	
DT	26-JUN-2001 (first entry)	
XX	26-JUN-2001 (first entry)	
DE	Human CDNA clone (3'-primer) SEQ ID NO:6064.	
XX	Human; primer: detection; diagnosis; antisense therapy; gene therapy; ss.	
KW	Human; primer: detection; diagnosis; antisense therapy; gene therapy; ss.	
XX	Human; primer: detection; diagnosis; antisense therapy; gene therapy; ss.	
OS	Homo sapiens.	
XX	Homo sapiens.	
PN	EP1074617-A2.	
XX	EP1074617-A2.	
PD	07-FEB-2001.	
XX	07-FEB-2001.	
XX	28-JUL-2000; 2000EP-0116126.	
PF	28-JUL-2000; 2000EP-0116126.	
XX	28-JUL-2000; 2000EP-0116126.	
PR	29-JUL-1999; 99JP-0248036.	
PR	27-AUG-1999; 99JP-0300253.	
PR	11-JAN-2000; 2000JP-0118776.	
PR	02-MAY-2000; 2000JP-0183767.	
PR	09-JUN-2000; 2000JP-0241899.	
XX	(HELI-) HELIX RES INST.	
PA	(HELI-) HELIX RES INST.	
XX	Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;	
PI	Ishii S, Sugiyama T, Wakamatsu A, Negai K, Otsuki T;	
XX	WPI; 2001-318749/34.	

```
XX primer sets for synthesizing polynucleotides, particularly the 5602
PT full-length cDNAs defined in the specification, and for the detection
PT and/or diagnosis of the abnormality of the proteins encoded by the
PT full-length cDNAs -
XX
XX Claim 3; SEQ ID 6064; 2537bp + CD ROM; English.
XX
XX The present invention describes primer sets for synthesizing 5602
CC full-length cDNAs defined in the specification. Where a primer set
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesizing polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.
XX
SQ Sequence 580 BP; 122 A; 156 C; 126 G; 175 T; 1 other;
XX
Query Match          92.0%; Score 18.4; DB 22; Length 580;
Best Local Similarity 95.0%; Pred. No. 5.9;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
OY 1 TGTACTCAGGGAGTGAGAA 20
   ||||| ||||| |||||
Db 524 TGTACTTAGGGAGTGAGAA 505
XX
RESULT 2
AAH15301
ID AAH15301 standard; cDNA; 2310 BP.
XX
AC AAH15301;
XX
XX 26-JUN-2001 (first entry)
XX
DE Human cDNA sequence SEQ ID NO:13453.
XX
XX Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX
OS Homo sapiens.
XX
PN EP1074617-A2.
XX
PD 07-FEB-2001.
XX
PF 28-JUL-2000; 2000EP-0116126.
XX
XX 29-JUL-1999; 99UP-0248036.
XX 27-AUG-1999; 99UP-0300253.
XX 11-JAN-2000; 2000JP-0118776.
XX 02-MAY-2000; 2000JP-0183767.
XX 09-JUN-2000; 2000JP-0241899.
XX
XX (HELI-) HELIX RES INST.
XX
XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
```

```
DR WP1; 2001-318749/34.
XX
XX primer sets for synthesizing polynucleotides, particularly the 5602
PT full-length cDNAs defined in the specification, and for the detection
PT and/or diagnosis of the abnormality of the proteins encoded by the
PT full-length cDNAs -
XX
XX Claim 8; SEQ ID 13453; 2537bp + CD ROM; English.
XX
XX The present invention describes primer sets for synthesizing 5602
CC full-length cDNAs defined in the specification. Where a primer set
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesizing polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.
XX
SQ Sequence 2310 BP; 716 A; 421 C; 448 G; 725 T; 0 other;
XX
Query Match          92.0%; Score 18.4; DB 22; Length 2310;
Best Local Similarity 95.0%; Pred. No. 7;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
OY 1 TGTACTCAGGGAGTGAGAA 20
   ||||| ||||| |||||
Db 1786 TGTACTTAGGGAGTGAGAA 1805
XX
RESULT 3
AAD31145
ID AAD31145 standard; cDNA; 3582 BP.
XX
AC AAD31145;
XX
XX 21-AUG-2002 (first entry)
XX
DE Human laminin gamma2 chain cDNA.
XX
XX Human laminin gamma2 chain cDNA.
XX
XX Cancer; cell proliferation; integrin receptor; signalling pathway;
XX gene therapy; human; laminin gamma2; ss.
XX
OS Homo sapiens.
XX
XX Key Location/Qualifiers
FH CDS 1..3582
FT /tag= a
FT /product= "laminin gamma2 chain"
XX
XX WO200230465-A2.
XX
XX 18-APR-2002.
XX
XX 12-OCT-2001; 2001WO-US32127.
XX
XX 12-OCT-2000; 2000US-239705P.
XX 24-OCT-2000; 2000US-242812P.
XX
XX (UNRP ) UNIV ROCHESTER.
PA
```

XX Land H, Deleu L;
PI
XX WPI: 2002-416838/44.
DR P-PSDB; AAB4712.
XX Reducing proliferation of cancer cell, by inhibiting ligand binding to
PT integrin receptor on cell, reducing integrin-integrin interaction,
PT receptor clustering interaction or integrin-non-integrin protein
PI interaction -
XX
XX Disclosure: Page 85-86; 148pp; English.
XX
XX The invention relates to a method of reducing proliferation of cancer
CC cells, comprising inhibiting ligand binding to integrin receptor on
CC cancer cells, where integrin receptor comprises integrin, reducing
CC integrin-integrin interaction, integrin receptor clustering interaction
CC or integrin-non-integrin protein interaction, reducing production of
CC integrin or ligand of integrin receptor by cancer cells, or
CC interfering with integrin-signalling pathway. The present sequence
CC is human laminin gamma2 chain cDNA.
XX
XX Sequence 3582 BP; 938 A; 872 C; 1026 G; 746 T; 0 other;
SQ
Query Match 92.0%; Score 18.4; DB 24; Length 3582;
Best Local Similarity 95.0%; Pred. No. 7.4;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
OY 1 TGTTCACGCGGATGAGAA 20
DB 1291 TGTTCACGCGGATGAGAA 1310
RESULT 4
AAC83732
ID AAC83732 standard; cDNA; 3620 BP.
XX
XX AAC83732;
AC
XX 02-MAR-2001 (first entry)
DE Human laminin 5 cDNA, SEQ ID NO: 31.
XX
XX Human laminin 5; vulnereary; antiulcer; antiinflammatory; antidiabetic;
KW cell adhesion promoter; wound healing; ulcers; burn; skin graft;
KM periodontitis; gingivitis; Type I diabetes; angiogenesis regulation; ss.
XX
XX Homo sapiens.
OS
XX WO200066731-A2.
PN
XX 09-NOV-2000.
PD
XX 28-APR-2000; 2000WO-US11459.
PE
XX 30-APR-1999; 99US-0131720.
PR 21-AUG-1999; 99US-0149738.
PR 24-SEP-1999; 99US-0155945.
XX
XX (BIOS-) BIOSTATUM INC.
PA
XX
XX Boutaud A;
PI
XX WPI: 2000-687538/67.
DR P-PSDB; AAB48471.
XX
XX Laminin 5-expressing cells, used to accelerate wound healing associated
PT with diabetic foot ulcers, venous ulcers, pressure sores, skin surgery,
PT burns, acute wounds and skin grafts -
XX
XX Claim 4; Page 204-209; 232pp; English.
PS
XX The present sequence encodes a laminin 5 chain polypeptide. Recombinant
CC

CC laminin 5-expressing cells are used to accelerate wound healing,
CC especially diabetic foot ulcers, venous ulcers, pressure sores, skin
CC surgery, burns, acute wounds, skin grafts, corneal ulcerations,
CC gastro-intestinal ulcers, periodontitis, and gingivitis. They are also
CC used to improve the biocompatibility of medical devices, and to promote
CC cell adhesion to a surface. They can be used for the ex vivo treatment
CC of Type I diabetes. Laminin can also be used to regulate angiogenesis.
CC The cell line produces and secretes recombinant heterotrimeric laminin,
CC whereas prior art cell lines have been created that produce but do not
CC secrete only one or two chain laminins.
XX
XX Sequence 3620 BP; 959 A; 862 C; 1040 G; 759 T; 0 other;
SQ
Query Match 92.0%; Score 18.4; DB 21; Length 3620;
Best Local Similarity 95.0%; Pred. No. 7.4;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
OY 1 TGTTCACGCGGATGAGAA 20
DB 1228 TGTTCACGCGGATGAGAA 1247
RESULT 5
AAC83731
ID AAC83731 standard; cDNA; 3720 BP.
XX
XX AAC83731;
AC
XX 02-MAR-2001 (first entry)
DE Human laminin 5 cDNA, SEQ ID NO: 29.
XX
XX Human laminin 5; vulnereary; antiulcer; antiinflammatory; antidiabetic;
KW cell adhesion promoter; wound healing; ulcers; burn; skin graft;
KM periodontitis; gingivitis; Type I diabetes; angiogenesis regulation; ss.
XX
XX Homo sapiens.
OS
XX WO200066731-A2.
PN
XX 09-NOV-2000.
PD
XX 28-APR-2000; 2000WO-US11459.
PE
XX 30-APR-1999; 99US-0131720.
PR 21-AUG-1999; 99US-0149738.
PR 24-SEP-1999; 99US-0155945.
XX
XX (BIOS-) BIOSTATUM INC.
PA
XX
XX Boutaud A;
PI
XX WPI: 2000-687538/67.
DR P-PSDB; AAB48470.
XX
XX Laminin 5-expressing cells, used to accelerate wound healing associated
PT with diabetic foot ulcers, venous ulcers, pressure sores, skin surgery,
PT burns, acute wounds and skin grafts -
XX
XX Claim 4; Page 195-200; 232pp; English.
PS
XX
XX The present sequence encodes a laminin 5 chain polypeptide. Recombinant
CC laminin 5-expressing cells are used to accelerate wound healing,
CC especially diabetic foot ulcers, venous ulcers, pressure sores, skin
CC surgery, burns, acute wounds, skin grafts, corneal ulcerations,
CC gastro-intestinal ulcers, periodontitis, and gingivitis. They are also
CC used to improve the biocompatibility of medical devices, and to promote
CC cell adhesion to a surface. They can be used for the ex vivo treatment
CC of Type I diabetes. Laminin can also be used to regulate angiogenesis.
CC The cell line produces and secretes recombinant heterotrimeric laminin,
CC whereas prior art cell lines have been created that produce but do not
CC secrete only one or two chain laminins.
XX

SO Sequence 3720 BP; 969 A; 900 C; 1067 G; 784 T; 0 other;

Query Match 92.0%; Score 18.4; DB 21; Length 3720;

Best Local Similarity 95.0%; Pred. No. 7.5;

Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 TGTACTCAGGCGATGAGAA 20

DB 1328 TGTATTTCAGGCGATGAGAA 1347

RESULT 6

AA13324

ID AA13324 standard; cDNA; 4316 BP.

XX AA13324;

DT 13-NOV-1996 (first entry)

DE Kallinin/laminin 5 gamma-2 chain alternative coding sequence.

XX Kallinin; laminin; epidermolysis bullosa; junctional; probe;

KM detection; inhibit; monitor; malignancy; ss.

XX Homo sapiens.

OS Key Location/Qualifiers

FT CDS 118..3453

FT /tag= b

FT /note= "Kallinin/laminin 5 gamma-2 chain"

PN WO9610646-A1.

PD 11-APR-1996.

PF - 04-OCT-1995; 95WO-EP03918.

PR 04-OCT-1994; 94US-0317450.

PA (TRYG/) TRYGVASON K.

PI Kallunki P, Pyke C, Trygvason K;

DR WPI; 1996-209366/21.

DR P-PSDB; AAR91428.

PT Detection of Kallinin or laminin 5 expression in cells - useful to

PT detect, monitor and inhibit the invasive growth of cell in tissue,

PS Disclosure; Fig 4B; 37pp; English.

CC The present sequence is cDNA (Genbank Z15009), with an alternative

CC 3' end sequence from T13324, encoding the Kallinin/laminin 5 gamma-2

CC chain (R91428). The gamma-2 chain is of importance to patients suffering

CC from epidermolysis bullosa, esp. the junctional form (JEB). Probes and

CC antisense gamma-2 sequences derived from this sequence can be used to

CC detect, monitor and inhibit the invasive growth of cells in tissue,

CC partic. malignant tissue.

SO Sequence 4316 BP; 1158 A; 1033 C; 1226 G; 899 T; 0 other;

OY 1 TGTACTCAGGCGATGAGAA 20

DB 1408 TGTATTTCAGGCGATGAGAA 1427

RESULT 7

AA142911

ID AA142911 standard; cDNA; 4316 BP.

XX AA142911;

XX 05-AUG-2002 (first entry)

DE Laminin gamma-2 chain cDNA sequence 2.

XX Laminin gamma-2; gene; ss; cancer; laminin gamma-2 chain inhibition;

KM carcinoma inhibition; anti-gamma-2 chain antibody;

XX epithelial cell adhesion; laminin-5.

OS Unidentified.

XX Key Location/Qualifiers

FT CDS 118..3453

FT /tag= a

FT /product= "Laminin gamma-2 chain 2"

FT sig_peptide 118..183

FT mat_peptide 184..3450

FT /tag= c

FT /note= "Mature laminin gamma-2 chain 2"

PN US2002052307-A1.

PD 02-MAY-2002.

PF 08-JAN-2001; 2001US-0756071.

PR 07-JAN-2000; 2000US-175005P.

PR 04-OCT-1994; 94US-0317450.

PR 18-FEB-1997; 97US-0800593.

PR 15-SEP-2000; 2000US-0663147.

PA (TRYG/) TRYGVASON K.

PI (KALL/) KALLUNKI P.

PA (PYKE/) PYKE C.

PI Trygvason K, Kallunki P, Pyke C;

DR WPI; 2002-434824/46.

DR P-PSDB; AA014993.

PT Modulating laminin 5 gamma 2 chain interactions of invasive carcinogens

PT for treating cancers and promoting attachment of cultured cells in

PT vitro -

PS Example 1; Page 26-30; 51pp; English.

CC The invention comprises a method of inhibiting the laminin gamma-2 chain

CC interactions of invasive carcinogens with surrounding tissues - by using

CC anti-gamma-2 chain antibodies to inhibit the gamma-2 chain biological

CC activity of the invasive carcinogens. The invention also comprises a

CC method for promoting adhesion of epithelial cells by exposing the cells

CC to intact laminin-5 molecules. The first method of the invention is

CC useful for preventing gamma 2 chain interactions of invasive carcinogens

CC with surrounding tissues. The second method of the invention is useful

CC for promoting adhesion of cultured epithelial cells. The present cDNA

CC sequence encodes a laminin gamma-2 chain.

SO Sequence 4316 BP; 1158 A; 1033 C; 1226 G; 899 T; 0 other;

OY 1 TGTACTCAGGCGATGAGAA 20

DB 1408 TGTATTTCAGGCGATGAGAA 1427

RESULT 8


```
XX Wang T, Fan L;
PI MPI: 2000-628399/60.
XX
XX Isolated polypeptide comprising an immunogenic portion of a lung tumor
PT protein is used for detecting and monitoring progression of lung cancer
PT in a patient -
XX
XX Claim 25a; Page 152-153; 261pp; English.
XX
XX This invention describes a novel isolated polypeptide (I) which
XX comprising an immunogenic portion of a lung tumor protein or variant (P2)
XX which have cytostatic activity. The polypeptides and polynucleotides are
XX used in compositions and vaccines to inhibit the development of cancer,
XX especially lung cancer, in a patient. Methods described in the invention
XX can be used to monitor the progression of a cancer by carrying out the
XX detection at subsequent time points and comparing the results from the
XX different time points. CD4+ and/or CD8+ T-Cells isolated from a patient
XX are treated with P2, polynucleotides encoding P2 or antigen presenting
XX cells expressing P2 and then administered to the patient to inhibit
XX development of cancer.
XX
XX Sequence 5156 BP; 1351 A; 1222 C; 1377 G; 1206 T; 0 other;
SQ
XX
XX Query Match          92.0%; Score 18.4; DB 21; Length 5156;
XX Best Local Similarity 95.0%; Pred. No. 7.8;
XX Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 TGTACTCAGGAGATGAGAA 20
DB 1356 TGTATTCAGGGAGTACGAA 1375
XX
XX RESULT 11
XX ABL49085
XX ID ABL49085 standard; cDNA; 5156 BP.
XX
XX ABL49085;
XX
XX 01-MAY-2002 (first entry)
XX
XX Human lung tumour cDNA sequence for contig 20 SEQ ID NO:130.
XX
XX Human; lung tumour; lung cancer; cytostatic; immunostimulant; vaccine;
XX immune response; ss.
XX
XX Homo sapiens.
XX
XX WO200200174-A2.
XX
XX 03-JAN-2002.
XX
XX
XX 28-JUN-2001; 2001WO-US21065.
XX
XX
XX 28-JUN-2000; 2000US-0606421.
XX
XX 02-AUG-2000; 2000US-0630940.
XX
XX 21-AUG-2000; 2000US-0643597.
XX
XX 15-SEP-2000; 2000US-0662786.
XX
XX 09-OCT-2000; 2000US-0685696.
XX
XX 12-DEC-2000; 2000US-0735705.
XX
XX 07-MAY-2001; 2001US-0850716.
XX
XX
XX (CORI-) CORIXA CORP.
XX
XX Wang T, Wang A, Skelky YAM, Li SX, Kalos MD, Henderson RA;
XX McNeill PD, Fanger N, Retter MW, Marnierakis M, Fanger GR;
XX Vedrick TS, Carter D, Watanabe Y, Peckham DW;
XX
XX MPI: 2002-090513/12.
XX
XX Polynucleotides encoding lung tumor polypeptides, useful for treating
XX lung cancer or stimulating an immune response -
PT
```

```
XX Example 3; Page 236-237; 374pp; English.
XX
XX The present invention describes human lung tumour proteins. Human lung
XX tumour proteins and polynucleotides have cytostatic and immunostimulant
XX activities, and can be used in vaccine production. Compositions
XX comprising the lung tumour proteins, polynucleotides, antibodies,
XX fusion proteins, T cell populations, or antigen presenting cells that
XX express the lung tumour proteins are useful for treating lung cancer or
XX stimulating an immune response. ABL49085 to ABL49300 and ABL4946 to
XX ABL75070 represent sequences used in the exemplification of the present
XX invention.
XX
XX Sequence 5156 BP; 1351 A; 1222 C; 1377 G; 1206 T; 0 other;
SQ
XX
XX Query Match          92.0%; Score 18.4; DB 24; Length 5156;
XX Best Local Similarity 95.0%; Pred. No. 7.8;
XX Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 TGTACTCAGGAGATGAGAA 20
DB 1356 TGTATTCAGGGAGTACGAA 1375
XX
XX RESULT 12
XX AAT13323
XX ID AAT13323 standard; cDNA; 5200 BP.
XX
XX AAT13323;
XX
XX 13-NOV-1996 (first entry)
XX
XX Kalinin/Laminin 5 gamma-2 chain coding sequence.
XX
XX Kalinin/Laminin 5 gamma-2 chain coding sequence.
XX
XX Kalinin; laminin; epidermolysis bullosa; junctional; probe;
XX detection; inhibit; monitor; malignancy; ss.
XX
XX Homo sapiens.
XX
XX Key Location/Qualifiers
XX CDS 118..3699
XX FT /tag=a
XX FT /product= kalinin/Laminin-5-gamma-2_chain
XX
XX WO9610646-A1.
XX
XX 11-APR-1996.
XX
XX 04-OCT-1995; 95WO-EP03918.
XX
XX 04-OCT-1994; 94US-0317450.
XX
XX (TRYG/) TRYGGVASON K.
XX
XX kallunk1 P, Pyke C, Tryggvason K;
XX
XX MPI: 1996-209366/21.
XX
XX P-PSDB; AAR91427.
XX
XX Detection of kalinin or laminin 5 expression in cells - useful to
XX detect, monitor and inhibit the invasive growth of cell in tissue,
XX partic. malignant tissue
XX
XX Disclosure; Fig 4A; 37pp; English.
XX
XX The present sequence is the full cDNA (Genbank Z15008) encoding the
XX kalinin/laminin 5 gamma-2 chain (R91427). The gamma-2 chain is of
XX importance to patients suffering from epidermolysis bullosa, esp. the
XX junctional form (JEB). Probes and antisense gamma-2 sequences derived
XX from this sequence can be used to detect, monitor and inhibit the
XX invasive growth of cells in tissue, partic. malignant tissue.
XX
XX Sequence 5200 BP; 1364 A; 1236 C; 1392 G; 1208 T; 0 other;
SQ
```


Query Match 92.0%; Score 18.4; DB 17; Length 5200;
Best Local Similarity 95.0%; Pred. No. 7.8;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 TGTACTCAGGCGATGAGAA 20
DB 1408 TGTATTTCAGGCGATGAGAA 1427

RESULT 13

AAC83729
ID AAC83729 standard; cDNA; 5200 BP.

AC AAC83729;

DT 02-MAR-2001 (first entry)

DE Human laminin 5 cDNA, SEQ ID NO: 25.

KM Human; laminin 5; vulnereary; antilucer; antinflammatory; antidiabetic;
cell adhesion promoter; wound healing; ulcers; burn; skin graft;

KM periodontitis; gingivitis; Type I diabetes; angiogenesis regulation; ss.

OS Homo sapiens.

PN WO200066731-A2.

PD 09-NOV-2000.

PF 28-APR-2000; 2000WO-US11459.

PR 30-APR-1999; 99US-0131720.

PR 21-AUG-1999; 99US-0149738.

PR 24-SEP-1999; 99US-0155945.

PA (BIOS-) BIOSTATUM INC.

PI Boutaud A;

DR WPI; 2000-687538/67.

DR P-PSDB; AAB48468.

PT Laminin 5-expressing cells, used to accelerate wound healing associated
with diabetic foot ulcers, venous ulcers, pressure sores, skin surgery,
burns, acute wounds and skin grafts -

PS Claim 4; Page 175-181; 232pp; English.

CC The present sequence encodes a laminin 5 chain polypeptide. Recombinant
laminin 5-expressing cells are used to accelerate wound healing,
especially diabetic foot ulcers, venous ulcers, pressure sores, skin
surgery, burns, acute wounds, skin grafts, corneal ulcerations,
gastro-intestinal ulcers, periodontitis, and gingivitis. They are also
used to improve the biocompatibility of medical devices, and to promote
cell adhesion to a surface. They can be used for the ex vivo treatment
of Type I diabetes. Laminin can also be used to regulate angiogenesis.
The cell line produces and secretes recombinant heterotrimeric laminin,
whereas prior art cell lines have been created that produce but do not
secrete only one or two chain laminins.

CC Sequence 5200 BP; 1364 A; 1236 C; 1392 G; 1208 T; 0 other;

Query Match 92.0%; Score 18.4; DB 21; Length 5200;

Best Local Similarity 95.0%; Pred. No. 7.8;

Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 TGTACTCAGGCGATGAGAA 20
DB 1408 TGTATTTCAGGCGATGAGAA 1427

RESULT 14

AAL42910
ID AAL42910 standard; cDNA; 5200 BP.

AC AAL42910;

DT 05-AUG-2002 (first entry)

DE Laminin gamma-2 chain cDNA sequence 1.

KM laminin gamma-2; gene; ss; cancer; laminin gamma-2 chain inhibition;
carcinogen inhibition; anti-gamma-2 chain antibody;

KM epithelial cell adhesion; laminin-5.

OS Unidentified.

PH Key Location/Qualifiers

FT CDS 118..3699

FT /tag= a

FT /product= "Laminin gamma-2 chain 1"

FT sig_peptide 118..183

FT mat_peptide 184..3696

FT /tag= b

FT /note= "Mature laminin gamma-2 chain 1"

PN US2002052307-A1.

PD 02-MAY-2002.

PF 08-JAN-2001; 2001US-0756071.

PR 07-JAN-2000; 2000US-175005P.

PR 04-OCT-1994; 94US-0317450.

PR 18-FEB-1997; 97US-0800593.

PR 15-SEP-2000; 2000US-0663147.

PA (TRYG/) TRYGIVASON K.

PA (KALL/) KALLUNKI P.

PA (PYKE/) PYKE C.

PI Tryggevason K, Kallunki P, Pyke C;

DR WPI; 2002-434824/46.

DR P-PSDB; AAO14992.

PT Modulating laminin 5 gamma 2 chain interactions of invasive carcinogens
for treating cancers and promoting attachment of cultured cells in
vitro -

PS Example 1; Fig 4; 51pp; English.

CC The invention comprises a method of inhibiting the laminin gamma-2 chain
interactions of invasive carcinogens with surrounding tissues - by using
anti-gamma-2 chain antibodies to inhibit the gamma-2 chain biological
activity of the invasive carcinogens. The invention also comprises a
method for promoting adhesion of epithelial cells by exposing the cells
to intact laminin-5 molecules. The first method of the invention is
useful for preventing gamma 2 chain interactions of invasive carcinogens
with surrounding tissues. The second method of the invention is useful
for promoting adhesion of cultured epithelial cells. The present cDNA
sequence encodes a laminin gamma-2 chain.

CC Sequence 5200 BP; 1364 A; 1236 C; 1392 G; 1208 T; 0 other;

Query Match 92.0%; Score 18.4; DB 24; Length 5200;

Best Local Similarity 95.0%; Pred. No. 7.8;

Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 TGTACTCAGGCGATGAGAA 20
DB 1408 TGTATTTCAGGCGATGAGAA 1427

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 1, 2003, 06:21:32 : Search time 26.6667 seconds
(Without alignments)
230.008 Million cell updates/sec

Title: US-10-053-662a-29

Perfect score: 20

Sequence: 1 tftactcagggagatgagaa 20

Scoring table: IDENTITY_NUC

Searched: Gapop 10.0 , Gapext 1.0

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

1: Issued_Patents_NM:*
2: /cgn2_6/ptodata/1/1na/5A_COMB.seq:*
3: /cgn2_6/ptodata/1/1na/5B_COMB.seq:*
4: /cgn2_6/ptodata/1/1na/6A_COMB.seq:*
5: /cgn2_6/ptodata/1/1na/6B_COMB.seq:*
6: /cgn2_6/ptodata/1/1na/PCrUS_COMB.seq:*
7: /cgn2_6/ptodata/1/1na/backfile1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	18.4	92.0	4316	1 US-08-317-450B-14	Sequence 14, Appl
2	18.4	92.0	4316	3 US-08-800-593-14	Sequence 14, Appl
3	18.4	92.0	5156	4 US-09-643-597-130	Sequence 130, App
4	18.4	92.0	5200	1 US-08-317-450B-12	Sequence 12, Appl
5	18.4	92.0	5200	3 US-08-800-593-12	Sequence 12, Appl
6	15.2	76.0	454	4 US-09-381-150A-1	Sequence 1, Appl
7	15.2	76.0	696	4 US-09-381-150A-2	Sequence 1, Appl
8	15.2	76.0	1810	1 US-07-913-107-1	Sequence 1, Appl
9	15.2	76.0	1810	1 US-08-459-201-1	Sequence 1, Appl
10	15.2	76.0	1810	1 US-08-281-248-1	Sequence 1, Appl
11	15.2	76.0	19227	3 US-09-090-793-13	Sequence 13, Appl
12	15.2	76.0	40138	3 US-09-090-793-12	Sequence 12, Appl
13	15.2	76.0	176373	3 US-09-128-155-17	Sequence 17, Appl
14	14.8	74.0	300	1 US-08-765-081-12	Sequence 12, Appl
15	14.8	74.0	300	3 US-09-098-082-12	Sequence 12, Appl
16	14.8	74.0	300	5 PCT-US95-06994-14	Sequence 14, Appl
17	14.8	74.0	485	4 US-09-484-970B-120	Sequence 120, App
18	14.8	74.0	2589	1 US-08-653-740-6	Sequence 6, Appl
19	14.8	74.0	2589	2 US-09-073-594-6	Sequence 6, Appl
20	14.8	74.0	2589	3 US-09-275-925-6	Sequence 6, Appl
21	14.8	74.0	17949	4 US-09-087-465-3	Sequence 3, Appl
22	14.8	74.0	43360	4 US-09-453-702B-206	Sequence 206, App
23	14.8	74.0	45325	4 US-09-453-702B-261	Sequence 261, App
24	14.8	74.0	169998	4 US-09-676-610B-24	Sequence 24, Appl
25	14.4	72.0	673	4 US-09-328-111-303	Sequence 303, App
26	14.4	72.0	1322	2 US-08-805-965-2	Sequence 2, Appl
27	14.4	72.0	1502	2 US-08-805-965-6	Sequence 6, Appl

ALIGNMENTS

28	14.4	72.0	2175	4 US-08-914-999-3	Sequence 3, Appl
29	14.2	71.0	263	3 US-09-004-113-8	Sequence 8, Appl
30	14.2	71.0	541	4 US-09-404-879A-11	Sequence 11, Appl
31	14.2	71.0	773	4 US-09-227-357-47	Sequence 47, Appl
32	14.2	71.0	1001	4 US-09-641-638-465	Sequence 465, App
33	14.2	71.0	1001	4 US-09-641-638-466	Sequence 466, App
34	14.2	71.0	1002	4 US-09-641-638-595	Sequence 595, App
35	14.2	71.0	1491	1 US-08-137-614A-3	Sequence 3, Appl
36	14.2	71.0	1491	2 US-08-768-301-1	Sequence 1, Appl
37	14.2	71.0	1701	4 US-09-411-628-5	Sequence 5, Appl
38	14.2	71.0	1861	4 US-09-129-668-7	Sequence 7, Appl
39	14.2	71.0	1897	1 US-08-184-632-1	Sequence 1, Appl
40	14.2	71.0	1981	4 US-08-981-392-26	Sequence 26, Appl
41	14.2	71.0	2065	4 US-09-129-668-5	Sequence 5, Appl
42	14.2	71.0	2178	3 US-09-034-916-1	Sequence 1, Appl
43	14.2	71.0	2265	3 US-08-906-865-2	Sequence 2, Appl
44	14.2	71.0	2265	4 US-09-129-668-2	Sequence 2, Appl
45	14.2	71.0	2457	4 US-08-872-757-1	Sequence 1, Appl

RESULT 1
US-08-317-450B-14
Sequence 14, Application US/08317450B
Patent No. 5660982
GENERAL INFORMATION:
APPLICANT: Trygvason, Karl
APPLICANT: Kallunki, Pekka
APPLICANT: Pyke, Charles
TITLE OF INVENTION: Laminin Chains: Diagnostic and
TITLE OF INVENTION: Therapeutic Use
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: BANNER & ALLEGRETTI, LTD.
STREET: Ten South Wacker Drive
CITY: Chicago
STATE: Illinois
COUNTRY: USA
ZIP: 60606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/317,450B
FILING DATE: 04-OCT-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Chao, Mark
REGISTRATION NUMBER: 37,293
REFERENCE/DOCKET NUMBER: 94,778
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312-715-1000
TELEFAX: 312-715-1234
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 4316 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: sig_peptide
LOCATION: 118..183
FEATURE:
NAME/KEY: CDS
LOCATION: 118..3453
FEATURE:
NAME/KEY: repeat_unit
LOCATION: 4021..4316

```

; OTHER INFORMATION: /rpt_type="other"
; OTHER INFORMATION: /rpt_family="HUMAN ALU"
; FEATURE:
; NAME/KEY: polyA_site
; LOCATION: 4296
; US-08-317-450B-14
;
Query Match          92.0%; Score 18.4; DB 1; Length 4316;
Best Local Similarity 95.0%; Pred. No. 0.64;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 TGTACTCAGGGGATGAGAA 20
        ||||| ||||| ||||| |||||
DB      1408 TGTATTCAGGGGATGAGAA 1427

RESULT 2
US-08-800-593-14
; Sequence 14, Application US/08800593
; Patent No. 6143505
; GENERAL INFORMATION:
; APPLICANT: Trygvason, Karl
; APPLICANT: Kallunki, Pekka
; APPLICANT: Pyke, Charles
; TITLE OF INVENTION: Laminin Chains: Diagnostic and
; NUMBER OF INVENTION: Therapeutic Use
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: McDonnell Boehnen Hulbert & Berghoff
; STREET: 300 South Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/800,593
; FILING DATE: 18-FEB-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/317,450
; FILING DATE: 04-OCT-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Chao, Mark
; REGISTRATION NUMBER: 37,293
; REFERENCE/DOCKET NUMBER: 94,778-B
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312-913-0001
; TELEFAX: 312-913-0002
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4316 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: sig_peptide
; LOCATION: 118..183
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 118..3453
; FEATURE:
; NAME/KEY: repeat_unit
; LOCATION: 4021..4316
; OTHER INFORMATION: /rpt_type="other"
; OTHER INFORMATION: /rpt_family="HUMAN ALU"
; FEATURE:
```

```

; NAME/KEY: polyA_site
; LOCATION: 4296
; US-08-800-593-14
;
Query Match          92.0%; Score 18.4; DB 3; Length 4316;
Best Local Similarity 95.0%; Pred. No. 0.64;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 TGTACTCAGGGGATGAGAA 20
        ||||| ||||| ||||| |||||
DB      1408 TGTATTCAGGGGATGAGAA 1427

RESULT 3
US-09-643-597-130
; Sequence 130, Application US/09643597
; Patent No. 6426072
; GENERAL INFORMATION:
; APPLICANT: Wang, Tongtong
; APPLICANT: Fan, Liqun
; APPLICANT: Kalos, Michael D.
; APPLICANT: Bangur, Chaitanya S.
; APPLICANT: Hosken, Nancy
; APPLICANT: Fanger, Gary R.
; APPLICANT: Li, Samuel X.
; APPLICANT: Wang, Aljun
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Henderson, Robert A.
; APPLICANT: McNeill, Patricia D.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.455C11
; CURRENT APPLICATION NUMBER: US/09/643,597
; CURRENT FILING DATE: 2000-08-21
; NUMBER OF SEQ ID NOS: 369
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 130
; LENGTH: 5156
; TYPE: DNA
; ORGANISM: Homo sapien
; US-09-643-597-130
;
Query Match          92.0%; Score 18.4; DB 4; Length 5156;
Best Local Similarity 95.0%; Pred. No. 0.66;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 TGTACTCAGGGGATGAGAA 20
        ||||| ||||| ||||| |||||
DB      1356 TGTATTCAGGGGATGAGAA 1375

RESULT 4
US-08-317-450B-12
; Sequence 12, Application US/08317450B
; Patent No. 5660962
; GENERAL INFORMATION:
; APPLICANT: Trygvason, Karl
; APPLICANT: Kallunki, Pekka
; APPLICANT: Pyke, Charles
; TITLE OF INVENTION: Laminin Chains: Diagnostic and
; NUMBER OF INVENTION: Therapeutic Use
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BANNER & ALLEGRETTI, LTD.
; STREET: Ten South Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
```

SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA: US/08/317,4508
APPLICATION NUMBER: US/08/317,4508
FILING DATE: 04-OCT-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Chao, Mark
REGISTRATION NUMBER: 37,293
REFERENCE/DOCKET NUMBER: 94,778
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312-715-1000
TELEFAX: 312-715-1234
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 5200 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
FEATURE:
NAME/KEY: sig_peptide
LOCATION: 118..183
FEATURE:
NAME/KEY: CDS
LOCATION: 118..3699
FEATURE:
NAME/KEY: polyA_site
LOCATION: 4433
FEATURE:
NAME/KEY: polyA_site
LOCATION: 5195
US-08-317-4508-12

Query Match 92.0%; Score 18.4; DB 1; Length 5200;
Best Local Similarity 95.0%; Pred. No. 0.66;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TGTTCAGGGGATGAGAA 20
DB 1408 TGTATTCAGGGGATGAGAA 1427

RESULT 5
US-08-800-593-12
Sequence 12, Application US/08800593
Patent No. 6143505
GENERAL INFORMATION:
APPLICANT: Tryggyvason, Karl
APPLICANT: Kallunki, Pekka
APPLICANT: Pyke, Charles
TITLE OF INVENTION: Laminin Chains: Diagnostic and
TITLE OF INVENTION: Therapeutic Use
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: McDonnell Boehnen Hulbert & Berghoff
STREET: 300 South Wacker Drive
CITY: Chicago
STATE: Illinois
COUNTRY: USA
ZIP: 60606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/800,593
FILING DATE: 18-FEB-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/317,450
FILING DATE: 04-OCT-1994
CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:
NAME: Chao, Mark
REGISTRATION NUMBER: 37,293
REFERENCE/DOCKET NUMBER: 94,778-B
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312-913-0001
TELEFAX: 312-913-0002
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 5200 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
FEATURE:
NAME/KEY: sig_peptide
LOCATION: 118..183
FEATURE:
NAME/KEY: CDS
LOCATION: 118..3699
FEATURE:
NAME/KEY: polyA_site
LOCATION: 4433
FEATURE:
NAME/KEY: polyA_site
LOCATION: 5195
US-08-800-593-12

Query Match 92.0%; Score 18.4; DB 3; Length 5200;
Best Local Similarity 95.0%; Pred. No. 0.66;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TGTTCAGGGGATGAGAA 20
DB 1408 TGTATTCAGGGGATGAGAA 1427

RESULT 6
US-09-381-150A-1/C
Sequence 1, Application US/09381150A
Patent No. 6465718
GENERAL INFORMATION:
APPLICANT: Inze, Dirk
APPLICANT: Segers, Gerda
APPLICANT: De Veylder, Lieven
APPLICANT: Mironov, Vladimir
TITLE OF INVENTION: METHOD AND MEANS FOR MODULATING
CELL GROWTH AND CONTROL
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Nixon Peabody LLP
STREET: 990 Stewart Avenue
CITY: Garden City
STATE: New York, New York
ZIP: 11530
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/381,150A
FILING DATE: 13-MAR-2000
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: EP PCT/EP98/01522
FILING DATE: 13-MAR-1998
APPLICATION NUMBER: EP 97.200.765.2
FILING DATE: 14-MAR-1997
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 454 base pairs

ADDRESSEE: DYKEMA GOSSETT
STREET: 400 Renaissance Center
CITY: Detroit
STATE: MI
COUNTRY: USA
ZIP: 48243
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/459,201
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/913,107
FILING DATE: 14-JUL-1992
ATTORNEY/AGENT INFORMATION:
NAME: Kelly, Robert L.
REGISTRATION NUMBER: 31,843
REFERENCE/DOCKET NUMBER: 61,686-016
TELECOMMUNICATION INFORMATION:
TELEPHONE: (313) 540-0849
TELEFAX: (313) 540-0763
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1810 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 177..1145
US-08-459-201-1

Query Match 76.0%; Score 15.2; DB 1; Length 1810;
Best Local Similarity 85.0%; Pred. No. 30;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TGTACTCAGGCGATGAGAA 20
DB 590 TGTTCCTAAGCGGAAGAGAA 571
||||| ||||||| |||||

RESULT 10
US-08-281-248-1/c
Sequence 1, Application US/08281248
Patent No. 5650500
GENERAL INFORMATION:
APPLICANT: Raz, Avraham
APPLICANT: Nabl, Ivan R.
APPLICANT: Otto, Thomas
APPLICANT: Watanabe, Hideomi
TITLE OF INVENTION: Method of Determining Metastatic
TITLE OF INVENTION: Potential of Tumor Cells
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: DYKEMA GOSSETT
STREET: 400 Renaissance Center
CITY: Detroit
STATE: MI
COUNTRY: USA
ZIP: 48243
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/281,248
FILING DATE: 27-JUL-1994

CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/913,107
FILING DATE: 14-JUL-1992
ATTORNEY/AGENT INFORMATION:
NAME: Kelly, Robert L.
REGISTRATION NUMBER: 31,843
REFERENCE/DOCKET NUMBER: 61,686-016
TELECOMMUNICATION INFORMATION:
TELEPHONE: (313) 540-0849
TELEFAX: (313) 540-0763
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1810 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 177..1145
US-08-281-248-1

Query Match 76.0%; Score 15.2; DB 1; Length 1810;
Best Local Similarity 85.0%; Pred. No. 30;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TGTACTCAGGCGATGAGAA 20
DB 590 TGTTCCTAAGCGGAAGAGAA 571
||||| ||||||| |||||

RESULT 11
US-09-090-793-13
Sequence 13, Application US/09090793
Patent No. 6140486
GENERAL INFORMATION:
APPLICANT: Calgene, LLC
TITLE OF INVENTION: Production of polyunsaturated fatty acids by expression
TITLE OF INVENTION: of polyketide-like synthesis genes in plants
FILE REFERENCE: CGNE.131.0105
CURRENT FILING DATE: 1998-06-04
EARLIER FILING DATE: 1997-06-04
NUMBER OF SEQ ID NOS: 66
SOFTWARE: Patentln Ver. 2.0
SEQ ID NO 13
LENGTH: 19227
TYPE: DNA
ORGANISM: Vibrio marinus
US-09-090-793-13

Query Match 76.0%; Score 15.2; DB 3; Length 19227;
Best Local Similarity 85.0%; Pred. No. 47;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TGTACTCAGGCGATGAGAA 20
DB 15171 TGTACTCAGGCGATGAGAA 15190
||||| ||||||| |||||

RESULT 12
US-09-090-793-12
Sequence 12, Application US/09090793
Patent No. 6140486
GENERAL INFORMATION:
APPLICANT: Calgene, LLC
TITLE OF INVENTION: Production of polyunsaturated fatty acids by expression
TITLE OF INVENTION: of polyketide-like synthesis genes in plants
FILE REFERENCE: CGNE.131.0105
CURRENT FILING DATE: 1998-06-04
APPLICATION NUMBER: US/09/090,793

EARLIER APPLICATION NUMBER: 60/048,650
EARLIER FILING DATE: 1997-06-04
NUMBER OF SEQ ID NOS: 66
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 12
LENGTH: 40138
TYPE: DNA
ORGANISM: Vibrio marinus
US-09-090-793-12

Query Match
Best local Similarity 85.0%; Score 15.2; DB 3; Length 40138;
Pred. No. 34;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TGTACTCAGGAGATGAGAA 20
||||| ||||| |||||
Db 32154 TGTACGTAGGAGATAGAA 32173

RESULT 13
US-09-128-155-17/C
Sequence 17, Application US/09128155
Patent No. 6117654
GENERAL INFORMATION:
APPLICANT: Pan, Yang
TITLE OF INVENTION: NOVEL MOLECULES OF TANGO-77 RELATED PROTEIN FAMILY
FILE REFERENCE: 09404/052001
CURRENT APPLICATION NUMBER: US/09/128,155
EARLIER FILING DATE: 1998-08-03
EARLIER APPLICATION NUMBER: US 60/091,650
EARLIER FILING DATE: 1998-07-02
EARLIER APPLICATION NUMBER: US 60/054,646
EARLIER FILING DATE: 1997-08-04
NUMBER OF SEQ ID NOS: 18
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 17
LENGTH: 176373
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)..(176373)
OTHER INFORMATION: n = A,T,C or G
US-09-128-155-17

Query Match
Best local Similarity 85.0%; Score 15.2; DB 3; Length 176373;
Pred. No. 70;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TGTACTCAGGAGATGAGAA 20
||||| ||||| |||||
Db 7658 TTTGACTCAGGAGATAGAA 7639

RESULT 14
US-08-765-081-12
Sequence 12, Application US/08765081
Patent No. 5798260
GENERAL INFORMATION:
APPLICANT: Tarr, P.T., Bilge, S.S., Besser, T.E., Vary Jr., J.C.
TITLE OF INVENTION: Escherichia coli 0157:H7 Epithelial Adhesin
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: Christensen, O'Connor, Johnson and Kindness
STREET: 2800 Pacific First Center, 1420 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98101-2347
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette-3.5 inch, 1.44MB storage
COMPUTER: IBM PC compatible/Pentium

OPERATING SYSTEM: MS-Windows 3.1
SOFTWARE: Word for Windows-6.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/765,081
FILING DATE: March 26, 1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/06994
FILING DATE: June 7, 1995
APPLICATION NUMBER: US 08/265,714
FILING DATE: June 24, 1994
ATTORNEY/AGENT INFORMATION:
NAME: Shelton, Dennis K.
REGISTRATION NUMBER: 26,997
REFERENCE/DOCKET NUMBER: CHOR-1-10286
TELECOMMUNICATION INFORMATION:
TELEPHONE: 1-206-682-8100; 1-206-224-0779 (direct)
TELEFAX: 1-206-224-0779
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 300 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
DESCRIPTION: Genomic DNA fragment described on page 10
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Escherichia coli 0157:H7
STRAIN: A5,F4,N11
US-08-765-081-12

Query Match
Best local Similarity 88.9%; Score 14.8; DB 1; Length 300;
Pred. No. 36;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 TTACTCAGGAGATGAGAA 20
||||| ||||| |||||
Db 66 TAACACGATGATGAGAA 83

RESULT 15
US-09-098-082-12
Sequence 12, Application US/09098082
Patent No. 6040421
GENERAL INFORMATION:
APPLICANT: Tarr, P.T., Bilge, S.S., Besser, T.E., Vary Jr., J.C.
TITLE OF INVENTION: Escherichia coli 0157:H7 Epithelial Adhesin
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: Christensen, O'Connor, Johnson and Kindness
STREET: 2800 Pacific First Center, 1420 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98101-2347
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette-3.5 inch, 1.44MB storage
COMPUTER: IBM PC compatible/Pentium II
OPERATING SYSTEM: MS-Windows 95
SOFTWARE: Word for Windows-6.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/098,082
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/765,081
FILING DATE: March 26, 1997
APPLICATION NUMBER: PCT/US95/06994
FILING DATE: June 7, 1995
APPLICATION NUMBER: US 08/265,714
FILING DATE: June 24, 1994
ATTORNEY/AGENT INFORMATION:

NAME: Shelness, Diana K.
REGISTRATION NUMBER: 35,356
REFERENCE/DOCKET NUMBER: CHOR-1-12402
TELECOMMUNICATION INFORMATION: 1-206-682-8100; 1-206-224-0735 (direct)
TELEFAX: 1-206-224-0779
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 300 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
DESCRIPTION: Genomic DNA fragment described on page 10
DESCRIPTION: of the specification
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Escherichia coli O157:H7
STRAIN: A5,F4,N11
US-09-098-082-12

Query Match	74.0%	Score 14.8	DB 3	Length 300
Best Local Similarity	88.9%	Pred. No. 36		
Matches	16	Conservative	0	Mismatches 2; Indels 0; Gaps 0
QY	3	TTACTCAGGGGATGAGAA	20	
db	66	TAATCTCAGTGGATGAGAA	83	

Search completed: July 1, 2003, 07:23:56
Job time : 28.6667 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 1, 2003, 06:30:26 ; Search time 67.1795 seconds
(without alignments)
441.935 Million cell updates/sec

Title: US-10-053-662a-29

Perfect score: 20

Sequence: 1 tgtactcagggatgagaa 20

Scoring table: IDENTITY_NUC

Searched: 1055720 seqs, 742224136 residues

Total number of hits satisfying chosen parameters: 2111440

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published_Applications_NA.*
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq.*
2: /cgn2_6/ptodata/1/pubpna/PCR_NEW_PUB.seq.*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq.*
4: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq.*
5: /cgn2_6/ptodata/1/pubpna/PCRUS_PUBCOMB.seq.*
6: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq.*
7: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq.*
8: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq.*
9: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq.*
10: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq.*
11: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq.*
12: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq.*
13: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq.*
14: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	18.4	92.0	4316	9 US-10-227-738-14	Sequence 14, Appl
2	18.4	92.0	4316	9 US-09-756-071B-14	Sequence 14, Appl
3	18.4	92.0	5156	10 US-09-735-705-130	Sequence 130, App
4	18.4	92.0	5156	10 US-09-850-716A-130	Sequence 130, App
5	18.4	92.0	5156	10 US-09-897-778-130	Sequence 130, App
6	18.4	92.0	5175	9 US-10-171-311-114	Sequence 114, App
7	18.4	92.0	5200	9 US-10-227-738-12	Sequence 12, Appl
8	18.4	92.0	5200	10 US-09-756-071B-12	Sequence 12, Appl
9	18.4	92.0	5460	12 US-10-044-090-558	Sequence 558, App
10	16.8	84.0	186957	9 US-10-185-770-3	Sequence 3, Appl
11	16	80.0	22400	9 US-10-091-504-1385	Sequence 1385, Ap
12	16	80.0	22400	10 US-09-764-869-1385	Sequence 1385, Ap
13	15.8	79.0	12258	10 US-09-801-574-58	Sequence 58, Appl
14	15.4	77.0	486	10 US-09-783-550-4920	Sequence 4920, Ap
15	15.2	76.0	303	10 US-09-764-877-291	Sequence 291, App
16	15.2	76.0	425	10 US-09-764-877-2589	Sequence 2589, App
17	15.2	76.0	430	10 US-09-783-590-10809	Sequence 10809, A
18	15.2	76.0	454	9 US-10-225-966-1	Sequence 1, Appl
19	15.2	76.0	454	10 US-09-938-342-1	Sequence 1, Appl

C 20	15.2	76.0	489	9 US-09-918-995-28552	Sequence 28552, A
C 21	15.2	76.0	525	9 US-09-918-995-15287	Sequence 15287, A
C 22	15.2	76.0	696	9 US-10-225-966-2	Sequence 2, Appl
C 23	15.2	76.0	696	10 US-09-938-342-2	Sequence 2, Appl
C 24	15.2	76.0	700	10 US-09-770-149-239	Sequence 239, App
C 25	15.2	76.0	1208	9 US-10-012-542-14	Sequence 14, Appl
C 26	15.2	76.0	1810	9 US-10-153-666-239	Sequence 239, App
C 27	15.2	76.0	1962	9 US-10-062-831-24	Sequence 24, Appl
C 28	15.2	76.0	2335	10 US-09-917-800A-1569	Sequence 1569, Ap
C 29	15.2	76.0	2898	9 US-10-153-666-355	Sequence 355, App
C 30	15.2	76.0	3296	9 US-10-153-668-357	Sequence 357, App
C 31	15.2	76.0	3398	9 US-10-198-846-10944	Sequence 10944, A
C 32	15.2	76.0	3589	9 US-09-764-891-9611	Sequence 9611, Ap
C 33	15.2	76.0	6057	9 US-10-331-061-79	Sequence 79, Appl
C 34	15.2	76.0	19227	9 US-10-331-061-13	Sequence 13, Appl
C 35	15.2	76.0	21000	9 US-09-975-123-11	Sequence 11, Appl
C 36	15.2	76.0	40138	9 US-10-331-061-12	Sequence 12, Appl
C 37	15.2	76.0	78056	9 US-10-109-551-1	Sequence 1, Appl
C 38	15.2	76.0	176373	9 US-10-095-407-17	Sequence 17, Appl
C 39	15	75.0	206	10 US-09-864-761-21659	Sequence 21659, A
C 40	15	75.0	442	10 US-09-864-761-4926	Sequence 4926, Ap
C 41	15	75.0	1825	9 US-10-037-270-313	Sequence 313, App
C 42	14.8	74.0	439	10 US-09-867-701-5313	Sequence 5313, Ap
C 43	14.8	74.0	443	9 US-09-918-995-6887	Sequence 6887, Ap
C 44	14.8	74.0	451	10 US-09-983-965-2432	Sequence 2432, Ap
C 45	14.8	74.0	502	10 US-09-783-590-450	Sequence 450, App

ALIGNMENTS

RESULT 1
US-10-227-738-14
; Publication 14, Application US/10227738
; Sequence 14, Appl
; GENERAL INFORMATION:
; APPLICANT: Trygvason, Karl
; Kallunki, Pekka
; Pyke, Charles
; TITLE OF INVENTION: Laminin Chains: Diagnostic and
; Therapeutic Use
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: McDonnell Boehnen Hulbert & Berghoff
; STREET: 300 South Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/227,738
; FILING DATE: 26-Aug-2002
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/800,593
; FILING DATE: 18-FEB-1997
; APPLICATION NUMBER: US 08/317,450
; FILING DATE: 04-OCT-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Chao, Mark
; REGISTRATION NUMBER: 37,293
; REFERENCE/DOCKET NUMBER: 94,778-B
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312-913-0001
; TELEFAX: 312-913-0002
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4316 base pairs

TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: sig_peptide
LOCATION: 118..183
FEATURE:
NAME/KEY: CDS
LOCATION: 118..3453
FEATURE:
NAME/KEY: repeat_unit
LOCATION: 4021..4316
OTHER INFORMATION: /rpt_type="other"
/rpt_family="HUMAN ALT"
FEATURE:
NAME/KEY: polyA_site
LOCATION: 4296
SEQUENCE DESCRIPTION: SEQ ID NO: 14:
US-10-227-738-14

Query Match 92.0%; Score 18.4; DB 9; Length 4316;
Best Local Similarity 95.0%; Pred. No. 3.8;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 TGTACTCAGGGGATGAGAA 20
||||| |||||||
DB 1408 TGTATTTCAGGGGATGAGAA 1427

RESULT 2
US-09-756-071B-14
Sequence 14, Application US/09756071B
Patent No. US20020052307A1
- GENERAL INFORMATION:
APPLICANT: Tryggevason, Karl
Kallunki, Pekka
Pyke, Charles
TITLE OF INVENTION: Laminin Chains: Diagnostic Uses
NUMBER OF SEQUENCES: 20
CORRESPONDENCE ADDRESS:
ADDRESSEE: Ray Sharpe Fagan Minnich & McKee
STREET: 1100 Superior Ave, Suite 700
CITY: Cleveland
STATE: Ohio
COUNTRY: USA
ZIP: 44114
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/756,071B
FILING DATE: 08-Jan-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 09/663,147
FILING DATE: 150-September 2000
ATTORNEY/AGENT INFORMATION:
NAME: Minnich, Richard, J
REGISTRATION NUMBER: 24,175
REFERENCE/DOCKET NUMBER: TRV 20014
TELECOMMUNICATION INFORMATION:
TELEPHONE: 216-861-5582
TELEFAX: 216-241-1666
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 4316 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA

FEATURE:
NAME/KEY: sig_peptide
LOCATION: 118..183
FEATURE:
NAME/KEY: CDS
LOCATION: 118..3453
FEATURE:
NAME/KEY: repeat_unit
LOCATION: 4021..4316
OTHER INFORMATION: /rpt_type="other"
/rpt_family="HUMAN ALT"
FEATURE:
NAME/KEY: polyA_site
LOCATION: 4296
SEQUENCE DESCRIPTION: SEQ ID NO: 14:
US-09-756-071B-14

Query Match 92.0%; Score 18.4; DB 10; Length 4316;
Best Local Similarity 95.0%; Pred. No. 3.8;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 TGTACTCAGGGGATGAGAA 20
||||| |||||||
DB 1408 TGTATTTCAGGGGATGAGAA 1427

RESULT 3
US-09-735-705-130
Sequence 130, Application US/09735705
Patent No. US20020052329A1
- GENERAL INFORMATION:
APPLICANT: Wang, Tonglong
APPLICANT: Fan, Liqun
APPLICANT: Kalos, Michael D.
APPLICANT: Bangur, Chaitanya S.
APPLICANT: Hosken, Nancy
APPLICANT: Fanger, Gary R.
APPLICANT: Wang, Samuel X.
APPLICANT: Wang, Aijun
APPLICANT: Skeiky, Yasir A.W.
APPLICANT: Henderson, Robert A.
APPLICANT: McNeill, Patricia D.
APPLICANT: Fanger, Neil
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
FILE REFERENCE: 210121.455C14
CURRENT APPLICATION NUMBER: US/09/735,705
CURRENT FILING DATE: 2000-12-12
NUMBER OF SEQ ID NOS: 419
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 130
LENGTH: 5156
TYPE: DNA
ORGANISM: Homo sapien
US-09-735-705-130

Query Match 92.0%; Score 18.4; DB 10; Length 5156;
Best Local Similarity 95.0%; Pred. No. 3.9;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 TGTACTCAGGGGATGAGAA 20
||||| |||||||
DB 1356 TGTATTTCAGGGGATGAGAA 1375

RESULT 4
US-09-850-716A-130
Sequence 130, Application US/09850716A
Patent No. US20020115139A1
- GENERAL INFORMATION:
APPLICANT: Kalos, Michael D.
APPLICANT: McNeill, Patricia D.
APPLICANT: Retter, Marc W.

```

; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 210121.455C15
; CURRENT APPLICATION NUMBER: US/09/850,716A
; CURRENT FILING DATE: 2001-05-07
; NUMBER OF SEQ ID NOS: 440
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 130
; LENGTH: 5156
; TYPE: DNA
; ORGANISM: Homo sapien
; US-09-850-716A-130

Query Match          92.0%; Score 18.4; DB 10; Length 5156;
Best Local Similarity 95.0%; Pred. No.3.9;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy      1 TGTACTCAGGCGATGAGAA 20
        ||||| ||||| ||||| |||||
Db      1356 TGTATTTCAGGCGATGAGAA 1375

RESULT 5
US-09-897-778-130
; Sequence 130, Application US/09897778
; Patent No. US20020147143A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Tongtong
; APPLICANT: Marnerakis, Margarita
; APPLICANT: Fanger, Gary R.
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Carter, Darrick
; APPLICANT: Watanabe, Yoshihiro
; APPLICANT: Henderson, Robert A.
; APPLICANT: Peckham, David W.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 210121.455C16
; CURRENT APPLICATION NUMBER: US/09/897,778
; CURRENT FILING DATE: 2001-06-28
; NUMBER OF SEQ ID NOS: 467
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 130
; LENGTH: 5156
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-09-897-778-130

Query Match          92.0%; Score 18.4; DB 10; Length 5156;
Best Local Similarity 95.0%; Pred. No.3.9;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy      1 TGTACTCAGGCGATGAGAA 20
        ||||| ||||| ||||| |||||
Db      1356 TGTATTTCAGGCGATGAGAA 1375

RESULT 6
US-10-171-311-114
; Sequence 114, Application US/10171311
; Publication No. US20030087270A1
; GENERAL INFORMATION:
; APPLICANT: Schlegel, Robert
; APPLICANT: Chen, Yan
; APPLICANT: Zhao, Xumei
; APPLICANT: Monahan, John
; APPLICANT: Kamatkar, Shubhangi
; APPLICANT: Glatt, Karen
; APPLICANT: Ganavathapu, Manjula
; APPLICANT: Hoerssh, Sebastian
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY
```

```

; TITLE OF INVENTION: OF CERVICAL CANCER
; FILE REFERENCE: MRI-035
; CURRENT APPLICATION NUMBER: US/10/171,311
; CURRENT FILING DATE: 2002-06-12
; PRIOR APPLICATION NUMBER: US 60/298,159
; PRIOR FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: US 60/298,155
; PRIOR FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: US 60/335,936
; PRIOR FILING DATE: 2001-11-14
; NUMBER OF SEQ ID NOS: 238
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 114
; LENGTH: 5175
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-10-171-311-114

Query Match          92.0%; Score 18.4; DB 9; Length 5175;
Best Local Similarity 95.0%; Pred. No.3.9;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy      1 TGTACTCAGGCGATGAGAA 20
        ||||| ||||| ||||| |||||
Db      1380 TGTATTTCAGGCGATGAGAA 1399

RESULT 7
US-10-227-738-12
; Sequence 12, Application US/10227738
; Publication No. US20030100529A1
; GENERAL INFORMATION:
; APPLICANT: Tryggvason, Karl
; APPLICANT: Kallunki, Pekka
; APPLICANT: Pyke, Charles
; TITLE OF INVENTION: Laminin Chains: Diagnostic and
; TITLE OF INVENTION: Therapeutic Use
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: McDonnell Boehnen Hulbert & Berghoff
; STREET: 300 South Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/227,738
; FILING DATE: 26-Aug-2002
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/800,593
; FILING DATE: 18-FEB-1997
; APPLICATION NUMBER: US 08/317,450
; FILING DATE: 04-OCT-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Chao, Mark
; REGISTRATION NUMBER: 37,293
; REFERENCE/DOCKET NUMBER: 94,778-B
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312-913-0001
; TELEFAX: 312-913-0002
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5200 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
```

```
FEATURE:
NAME/KEY: sig_peptide
LOCATION: 118..183
FEATURE:
NAME/KEY: CDS
LOCATION: 118..3699
FEATURE:
NAME/KEY: polyA_site
LOCATION: 4433
FEATURE:
NAME/KEY: polyA_site
LOCATION: 5195
SEQUENCE DESCRIPTION: SEQ ID NO: 12:
US-10-227-738-12

Query Match
Best Local Similarity 92.0%; Score 18.4; DB 9; Length 5200;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TGTACTCAGGGGATGAGAA 20
Db 1408 TGTATTTCAGGGGATGAGAA 1427

RESULT 8
US-09-756-071B-12
Sequence 12, Application US/09756071B
Patent No. US20020052307A1
GENERAL INFORMATION:
APPLICANT: Trygvsason, Karl
Kallunki, Pekka
Pyke, Charles
TITLE OF INVENTION: Laminin Chains: Diagnostic Uses
NUMBER OF SEQUENCES: 20
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fay Sharpe Fagan Minnich & McKee
STREET: 1100 Superior Ave, Suite 700
CITY: Cleveland
STATE: Ohio
COUNTRY: USA
ZIP: 44114
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/756,071B
FILING DATE: 08-Jan-2001
CLASSIFICATION: <unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 09/663,147
FILING DATE: 15-September 2000
ATTORNEY/AGENT INFORMATION:
NAME: Minnich, Richard, J
REGISTRATION NUMBER: 24,175
REFERENCE/DOCKET NUMBER: TRV 20014
TELECOMMUNICATION INFORMATION:
TELEPHONE: 216-861-5582
TELEFAX: 216-241-1666
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 5200 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: sig_peptide
LOCATION: 118..183
FEATURE:
NAME/KEY: CDS
LOCATION: 118..3699
```

```
FEATURE:
NAME/KEY: polyA_site
LOCATION: 4433
FEATURE:
NAME/KEY: polyA_site
LOCATION: 5195
SEQUENCE DESCRIPTION: SEQ ID NO: 12:
US-09-756-071B-12

Query Match
Best Local Similarity 92.0%; Score 18.4; DB 10; Length 5200;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TGTACTCAGGGGATGAGAA 20
Db 1408 TGTATTTCAGGGGATGAGAA 1427

RESULT 9
US-10-044-090-558
Sequence 558, Application US/10044090
Patent No. US20020137081A1
GENERAL INFORMATION:
APPLICANT: Olga Bandman
TITLE OF INVENTION: GENES DIFFERENTIALLY EXPRESSED IN VASCULAR TISSUE ACTIVATION
FILE REFERENCE: PA-0028 US
CURRENT APPLICATION NUMBER: US/10/044,090
CURRENT FILING DATE: 2002-01-09
NUMBER OF SEQ ID NOS: 850
SOFTWARE: PERL Program
SEQ ID NO 558
LENGTH: 5460
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc_feature
OTHER INFORMATION: Incyte ID No. US20020137081A1 149791.5
US-10-044-090-558

Query Match
Best Local Similarity 92.0%; Score 18.4; DB 12; Length 5460;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TGTACTCAGGGGATGAGAA 20
Db 1410 TGTATTTCAGGGGATGAGAA 1429

RESULT 10
US-10-185-770-3
Sequence 3, Application US/10185770
Publication No. US20030022217A1
GENERAL INFORMATION:
APPLICANT: CRECARDI, Toni et al.
TITLE OF INVENTION: ISOLATED HUMAN SECRETED PROTEINS, AND
TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN SECRETED PROTEINS, AND
FILE REFERENCE: CL0001247
CURRENT APPLICATION NUMBER: US/10/185,770
CURRENT FILING DATE: 2002-07-01
PRIOR APPLICATION NUMBER: 60/301,852
PRIOR FILING DATE: 2001-07-02
NUMBER OF SEQ ID NOS: 4
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 3
LENGTH: 186957
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)...(186957)
OTHER INFORMATION: n = A,T,C or G
US-10-185-770-3
```

Query Match 84.0%; Score 16.8; DB 9; Length 186957;
Best Local Similarity 90.0%; Pred. No. 31;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TGTACTCAGGGGATGAGAA 20
DB 98798 TTTTATTCAGGGGATGAGAA 98817

RESULT 11

US-10-091-504-1385/C
; Sequence 1385, Application US/10091504
; Publication No. US20030059908A1
; GENERAL INFORMATION:

APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC007C1
; CURRENT APPLICATION NUMBER: US/10/091,504
; CURRENT FILING DATE: 2002-03-07
; NUMBER OF SEQ ID NOS: 2442
; Prior Application removed - See File Wrapper or Palm
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1385
; LENGTH: 22400
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-091-504-1385

Query Match 80.0%; Score 16; DB 9; Length 22400;
Best Local Similarity 100.0%; Pred. No. 74;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 TTACTCAGGGGATGAG 18
DB 944 TTACTCAGGGGATGAG 929

RESULT 12

US-09-764-869-1385/C
; Sequence 1385, Application US/09764869
; Patent No. US20020061521A1
; GENERAL INFORMATION:

APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC007
; CURRENT APPLICATION NUMBER: US/09/764,869
; CURRENT FILING DATE: 2001-01-17
; Prior Application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 2442
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1385
; LENGTH: 22400
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-869-1385

Query Match 80.0%; Score 16; DB 10; Length 22400;
Best Local Similarity 100.0%; Pred. No. 74;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 TTACTCAGGGGATGAG 18
DB 944 TTACTCAGGGGATGAG 929

RESULT 13

US-09-801-574-58
; Sequence 58, Application US/09801574
; Patent No. US20020081592A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Peijiang Jeremy
; APPLICANT: Page, David C.

; TITLE OF INVENTION: Reproduction-Specific Genes
; FILE REFERENCE: 0399.2007-002
; CURRENT APPLICATION NUMBER: US/09/801,574
; CURRENT FILING DATE: 2001-03-07
; PRIOR APPLICATION NUMBER: 60/187,518
; PRIOR FILING DATE: 2000-03-07
; PRIOR APPLICATION NUMBER: 60/261,557
; PRIOR FILING DATE: 2001-01-12
; NUMBER OF SEQ ID NOS: 90
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 58
; LENGTH: 12258
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-801-574-58

Query Match 79.0%; Score 15.8; DB 10; Length 12258;
Best Local Similarity 89.5%; Pred. No. 92;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TGTACTCAGGGGATGAGA 19
DB 9217 TGTACTCAGGGGATGAGA 9235

RESULT 14

US-09-783-590-4920/C
; Sequence 4920, Application US/09783590
; Patent No. US20020110850A1
; GENERAL INFORMATION:

APPLICANT: Dillon, Patrick J.
; APPLICANT: Haseltine, William A.
; APPLICANT: Li, Haodong
; APPLICANT: Rosen, Craig A.
; APPLICANT: Ruben, Steven M.
; TITLE OF INVENTION: Human Genes, Sequences, and Expression Products 16.2
; FILE REFERENCE: PO-16.2C1
; CURRENT APPLICATION NUMBER: US/09/783,590
; CURRENT FILING DATE: 2000-02-15
; PRIOR APPLICATION NUMBER: 08/420,856
; PRIOR FILING DATE: 1995-04-12
; PRIOR APPLICATION NUMBER: 08/346,731
; PRIOR FILING DATE: 1994-11-21
; NUMBER OF SEQ ID NOS: 12485
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4920
; LENGTH: 486
; TYPE: DNA
; ORGANISM: Homo sapiens

FEATURE:
NAME/KEY: misc feature
LOCATION: (10)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (100)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (106)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (254)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (266)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (273)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (274)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (314)

```
? OTHER INFORMATION: n equals a,t,g, or c
? NAME/KEY: misc feature
? LOCATION: (315)
? OTHER INFORMATION: n equals a,t,g, or c
? NAME/KEY: misc feature
? LOCATION: (320)
? OTHER INFORMATION: n equals a,t,g, or c
? NAME/KEY: misc feature
? LOCATION: (321)
? OTHER INFORMATION: n equals a,t,g, or c
? NAME/KEY: misc feature
? LOCATION: (329)
? OTHER INFORMATION: n equals a,t,g, or c
? NAME/KEY: misc feature
? LOCATION: (331)
? OTHER INFORMATION: n equals a,t,g, or c
? NAME/KEY: misc feature
? LOCATION: (349)
? OTHER INFORMATION: n equals a,t,g, or c
? NAME/KEY: misc feature
? LOCATION: (350)
? OTHER INFORMATION: n equals a,t,g, or c
? NAME/KEY: misc feature
? LOCATION: (353)
? OTHER INFORMATION: n equals a,t,g, or c
? NAME/KEY: misc feature
? LOCATION: (354)
? OTHER INFORMATION: n equals a,t,g, or c
? NAME/KEY: misc feature
? LOCATION: (356)
? OTHER INFORMATION: n equals a,t,g, or c
? NAME/KEY: misc feature
? LOCATION: (358)
? OTHER INFORMATION: n equals a,t,g, or c
? NAME/KEY: misc feature
? LOCATION: (359)
? OTHER INFORMATION: n equals a,t,g, or c
? NAME/KEY: misc feature
? LOCATION: (366)
? OTHER INFORMATION: n equals a,t,g, or c
? NAME/KEY: misc feature
? LOCATION: (369)
? OTHER INFORMATION: n equals a,t,g, or c
? NAME/KEY: misc feature
? LOCATION: (370)
? OTHER INFORMATION: n equals a,t,g, or c
? NAME/KEY: misc feature
? LOCATION: (380)
? OTHER INFORMATION: n equals a,t,g, or c
? NAME/KEY: misc feature
? LOCATION: (381)
? OTHER INFORMATION: n equals a,t,g, or c
? NAME/KEY: misc feature
? LOCATION: (392)
? OTHER INFORMATION: n equals a,t,g, or c
? NAME/KEY: misc feature
? LOCATION: (396)
? OTHER INFORMATION: n equals a,t,g, or c
? NAME/KEY: misc feature
? LOCATION: (401)
? OTHER INFORMATION: n equals a,t,g, or c
? NAME/KEY: misc feature
? LOCATION: (403)
? OTHER INFORMATION: n equals a,t,g, or c
? NAME/KEY: misc feature
? LOCATION: (406)
? OTHER INFORMATION: n equals a,t,g, or c
? NAME/KEY: misc feature
? LOCATION: (410)
? OTHER INFORMATION: n equals a,t,g, or c
? NAME/KEY: misc feature
? LOCATION: (417)
? OTHER INFORMATION: n equals a,t,g, or c
```

```
? NAME/KEY: misc feature
? LOCATION: (424)
? OTHER INFORMATION: n equals a,t,g, or c
? NAME/KEY: misc feature
? LOCATION: (425)
? OTHER INFORMATION: n equals a,t,g, or c
? NAME/KEY: misc feature
? LOCATION: (427)
? OTHER INFORMATION: n equals a,t,g, or c
? NAME/KEY: misc feature
? LOCATION: (430)
? OTHER INFORMATION: n equals a,t,g, or c
? NAME/KEY: misc feature
? LOCATION: (432)
? OTHER INFORMATION: n equals a,t,g, or c
? NAME/KEY: misc feature
? LOCATION: (442)
? OTHER INFORMATION: n equals a,t,g, or c
? NAME/KEY: misc feature
? LOCATION: (443)
? OTHER INFORMATION: n equals a,t,g, or c
? NAME/KEY: misc feature
? LOCATION: (444)
? OTHER INFORMATION: n equals a,t,g, or c
? NAME/KEY: misc feature
? LOCATION: (449)
? OTHER INFORMATION: n equals a,t,g, or c
? NAME/KEY: misc feature
? LOCATION: (450)
? OTHER INFORMATION: n equals a,t,g, or c
? NAME/KEY: misc feature
? LOCATION: (455)
? OTHER INFORMATION: n equals a,t,g, or c
? NAME/KEY: misc feature
? LOCATION: (457)
? OTHER INFORMATION: n equals a,t,g, or c
? NAME/KEY: misc feature
? LOCATION: (459)
? OTHER INFORMATION: n equals a,t,g, or c
? NAME/KEY: misc feature
? LOCATION: (474)
? OTHER INFORMATION: n equals a,t,g, or c
? NAME/KEY: misc feature
? LOCATION: (475)
? OTHER INFORMATION: n equals a,t,g, or c
? NAME/KEY: misc feature
? LOCATION: (477)
? OTHER INFORMATION: n equals a,t,g, or c
? US-09-783-590-4920
```

```
Query Match 77.0%; Score 15.4; DB 10; Length 486;
Best Local Similarity 94.1%; Pred. No. 1.3e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
QY 4 TACTCAGGGATGAGAA 20
Db 243 TACTCTGGGATGAGAA 227
```

```
RESULT 15
US-09-764-877-291/c
; Sequence 291, Application US/09764877
; Patent No. US20020147140A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC005
; CURRENT APPLICATION NUMBER: US/09/764,877
; PRIOR APPLICATION DATA REMOVED - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 4031
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 291
```



```
; LENGTH: 303
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (205)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (267)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (293)
; OTHER INFORMATION: n equals a,t,g, or c
; US-09-764-877-291
```

```
Query Match          76.0%; Score 15.2; DB 10; Length 303;
Best Local Similarity 85.0%; Pred. No. 1.6e+02;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

```
QY      1 TGTTCACAGGGGATGAGAA 20
          ||||| ||||| |||||
Db      83 TGTTCACAGGGGATGAGAA 64
```

Search completed: July 1, 2003, 07:26:17
Job time : 68.1795 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 1, 2003, 03:57:39 ; Search time 904.615 Seconds
(without alignments)
358.063 Million cell updates/sec

Title: US-10-053-662A-29

Perfect score: 20

Sequence: 1 tgttaccagggagatgagaa 20

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 16154066 segs, 8097743376 residues 32308132

Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*
1: em_estba:*
2: em_esthum:*
3: em_estnu:*
4: em_estov:*
5: em_estro:*
6: em_estro:*
7: em_estro:*
8: em_estro:*
9: gb_est1:*
10: gb_est2:*
11: gb_est3:*
12: gb_est4:*
13: gb_est5:*
14: gb_est5:*
15: em_estfun:*
16: em_estfun:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vit:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match length	ID	Description
1	18.4	92.0	261 14	BQ325508 QV2-C1010
2	18.4	92.0	413 14	T70002 ycl19f04.r1
3	18.4	92.0	530 9	AL710603 DKFZP6860
4	18.4	92.0	580 9	AU145040 AU145040
5	18.4	92.0	610 9	AL710604 DKFZP6860
6	18.4	92.0	782 12	BG679991 BG679991

Result No.	Score	Query Match length	ID	Description
7	18.4	92.0	1007 12	BG749494
8	17.4	87.0	446 17	AQ880436
9	17.4	87.0	560 17	AQ481409
10	16.8	84.0	194 17	A2328393
11	16.8	84.0	242 9	AA055712
12	16.8	84.0	263 9	AV2227530
13	16.8	84.0	385 17	AQ264952
14	16.8	84.0	437 17	AQ002138
15	16.8	84.0	442 10	BE247528
16	16.8	84.0	512 17	A2169304
17	16.8	84.0	550 14	BQ320138
18	16.8	84.0	576 17	TA66812P
19	16.8	84.0	628 17	A2992702
20	16.8	84.0	822 12	BF300476
21	16.8	84.0	893 14	BQ420274
22	16.8	84.0	914 12	BE671760
23	16.8	84.0	1044 12	BE748740
24	16.6	83.0	958 17	CNS02A60
25	16.4	82.0	60 17	CGA350111
26	16.4	82.0	269 17	AQ103499
27	16.4	82.0	382 17	BH364764
28	16.4	82.0	385 17	A2734473
29	16.4	82.0	398 12	BG100888
30	16.4	82.0	342 12	BH554244
31	16.4	80.0	440 12	BG043081
32	16.4	80.0	468 10	BE554551
33	16.4	80.0	752 17	BH384164
34	15.8	79.0	172 9	AV093257
35	15.8	79.0	182 13	BH091644
36	15.8	79.0	210 17	AQ310289
37	15.8	79.0	229 14	F05097
38	15.8	79.0	263 17	A2942797
39	15.8	79.0	301 14	R58268
40	15.8	79.0	348 10	BE465362
41	15.8	79.0	351 17	B96116
42	15.8	79.0	378 14	W03415
43	15.8	79.0	390 17	AQ060420
44	15.8	79.0	410 14	D39211
45	15.8	79.0	411 17	AQ305163

ALIGNMENTS

RESULT 1
BQ325508/c 261 bp mRNA linear EST 17-MAY-2002
QV2-C10104-020301-611-d06 C10104 Homo sapiens CDNA, mRNA sequence.
BQ325508
BQ325508.1 GI:20939888
EST.
SOURCE
ORGANISM
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE
AUTHORS
Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Britones, M.R.,
Nagal, M.A., da Silva, W. Jr., Zago, M.B., Bordin, S., Costa, F.F.,
Goldman, G.H., Carvalho, A.F., Matsukuma, A., Bala, G.S., Simpson, D.H.,
Brunstein, A., deoliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare,
M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and
Simpson, A.J.
TITLE
Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
JOURNAL
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
COMMENT
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001

Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL
(<http://www.ludwig.org.br/scripts/gethtml2.pl?l1=GV26f2-GV2-C10104-020301-611-d066f3-2001-03-02&f4=1>)

Seq primer: puc 18 forward
High quality sequence start: 5
High quality sequence stop: 261.
Location/Qualifiers

FEATURES

source

1. 261
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="C10104"
/dev_stage="Adult"
/note="Organ: colon_lns; Vector: puc18; Site_1: SmaI;
Site_2: SmaI; A mini-library was made by cloning products
derived from ORESTES PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the pUC 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."

BASE COUNT

54 a 79 c 76 g 52 t

ORIGIN

Query Match 92.0%; Score 18.4; DB 14; Length 261;
Best Local Similarity 95.0%; Pred. No. 76;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY

1 TGTACTCAGGCGATGACAA 20
||||| |||||||||
211 TGTATTACGGCGATGACAA 192

Db

RESULT 2
LOCUS T70002 413 bp mRNA linear EST 23-FEB-1995
DEFINITION ycl9f04.r1 Stratagene lung (#937210) Homo sapiens cDNA clone
IMAGE:81151 5' similar to contains MER20 repetitive element;; mRNA
sequence.
T70002
T70002.1 GI:681150

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

COMMENT

CONTACT

WASHINGTON

UNIVERSITY

SCHOOL

OF MEDICINE

4444

FOREST PARK

Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: est@watson.wustl.edu

Insert Size: 936

High quality sequence stops: 331 Source: IMAGE Consortium, LNL This

clone is available royalty-free through LNL; contact the IMAGE

Consortium (info@image.llnl.gov) for further information.

Insert length: 936 Std Error: 0.00

Seq primer: M13Rpl

High quality sequence stop: 331.

Location/Qualifiers

1. 413

/organism="Homo sapiens"

/db_xref="GDB:484768"

/db_xref="taxon:9606"
/clone_image:81151"
/clone_lib="Stratagene lung (#937210)"
/sex="male"
/dev_stage="72 years"
/lab_host="SOLR cells (kanamycin resistant)"
/note="Organ: lung; Vector: pBluescript SK-; Site_1: EcoRI
; Site_2: XhoI; Cloned unidirectionally. Primer: Oligo
dT, normal lung. Average insert size: 1.0 kb; Uni-ZAP XR
Vector: ~5' adaptor sequence: 5' GAATTCGACGAG 3' ~3'
adaptor sequence: 5' CTCGAGTTTCTTTTCTTTT 3'"

BASE COUNT

121 a 83 c 104 g 102 t 3 others

ORIGIN

Query Match 92.0%; Score 18.4; DB 14; Length 413;
Best Local Similarity 95.0%; Pred. No. 89;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY

1 TGTACTCAGGCGATGACAA 20
||||| |||||||||
230 TGTACTTAGGCGATGACAA 249

Db

RESULT 3
LOCUS AL710603 530 bp mRNA linear EST 22-MAR-2002
DEFINITION DKFZP6860036.r1 686 (synonym: hlcc3) Homo sapiens cDNA clone
DKFZP6860036 5', mRNA sequence.
AL710603
AL710603.1 GI:19693958

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

MIPS

Am

Klopferspitz 18a D-82152 Martinsried, Germany

This is the 5' sequence of the clone insert

Clone from S. Wiemann, Molecular Genome Analysis, German Cancer

Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;

sequenced by EMBL (European Molecular Biology Laboratories),

Heidelberg/Germany) within the cDNA sequencing consortium of the

German Genome Project.

No sl sequence available.

This clone (DKFZP6860036) is available at the RZPD in Berlin.

Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059

Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.

Location/Qualifiers

1. 530

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="DKFZP6860036"

/clone_lib="686 (synonym: hlcc3)"

/tissue_type="human skeletal muscle"

/dev_stage="adult"

/lab_host="DH10B"

/note="Vector: pTriplex2; Site_1: SfiI; Site_2: SfiIb;
cDNA-collection"

BASE COUNT

168 a 109 c 145 g 108 t

ORIGIN

Query Match 92.0%; Score 18.4; DB 9; Length 530;
Best Local Similarity 95.0%; Pred. No. 97;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY

1 TGTACTCAGGCGATGACAA 20
||||| |||||||||
61 TGTACTTAGGCGATGACAA 80

Db

RESULT 4
LOCUS AUI45040/c 580 bp mRNA linear EST 05-AUG-2002
DEFINITION AUI45040 HEMBA1 Homo sapiens cDNA clone HEMBA1003720 3', mRNA
sequence.
ACCESSION AUI45040
VERSION AUI45040.1 GI:11006561
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
AUTHORS Ota,T., Nishikawa,T., Suzuki,Y., Ishii,S., Salto,K., Kawai,Y.,
Yamamoto,J., Wakamatsu,A., Ozawa,M., Nakamura,Y., Nagai,T., Sugano
S. and Isogai,T.).
HRI human cDNA project (Ota,T., Nishikawa,T., Suzuki,Y., Ishii,S.,
Salto,K., Kawai,Y., Yamamoto,J., Wakamatsu,A., Ozawa,M., Nakamura
Y., Nagai,T., Sugano,S., Isogai,T.)
JOURNAL Unpublished (2000)
COMMENT Contact: Takao Isogai
Genomics Laboratory
Helix Research Institute
1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
Tel: 81-438-52-3925
Fax: 81-438-52-3986
Email: genomics@hri.co.jp
HRI human cDNA project; 5'- & 3'-end one pass sequencing; Helix
Research Institute; cDNA library construction; Department of
Virology, Institute of Medical Science, University of Tokyo, and
Helix Research Institute.
FEATURES
Source
Location/Qualifiers
1..580
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="HEMBA1003720"
/clone_lib="HEMBA1"
/tissue_type="whole embryo, mainly head"
/dev_stage="embryo, 10 weeks"
/note="vector: pME18SFL3"
BASE COUNT 122 a 156 c 126 g 175 t 1 others
ORIGIN
Query Match 92.0%; Score 18.4; DB 9; Length 580;
Best Local Similarity 95.0%; Pred. No. 1e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
OY 1 TGTACTCAGGGGATGAGAA 20
||||| |||||||||
DB 524 TGTACTTAAAGGATGAGAA 505
RESULT 5
LOCUS AL710604 610 bp mRNA linear EST 22-MAR-2002
DEFINITION DKFZp6860046.t1.686 (synonym: hlc3) Homo sapiens cDNA clone
DKFZp6860046 5', mRNA sequence.
ACCESSION AL710604
VERSION AL710604.1 GI:19693959
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
AUTHORS Ansoorge,W., Winkler,U., Mewes,W., Well,B. and Wiemann,S.
TITLE EST (Ansoorge,W., Winkler,U., Mewes,W., Well,B. and Wiemann,S.)
JOURNAL Unpublished (1999)
COMMENT Contact: Ansoorge W
MIS
Am Klopferspitz 18a D-82152 Martinsried, Germany

This is the 5' sequence of the clone insert
Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
Sequenced by EMBL (European Molecular Biology Laboratories,
Heidelberg/Germany) within the cDNA sequencing consortium of the
German Genome Project.
No sl sequence available.
This clone (DKFZp6860046) is available at the RZPD in Berlin.
Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.
FEATURES
Source
Location/Qualifiers
1..610
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="DKFZp6860046"
/clone_lib="686 (synonym: hlc3)"
/tissue_type="human skeletal muscle"
/dev_stage="adult"
/lab_host="DH10B"
/note="vector: pTR1p1Ex2; Site_1: SfiI; Site_2: SfiIb;
cDNA-collection"
BASE COUNT 199 a 126 c 161 g 124 t
ORIGIN
Query Match 92.0%; Score 18.4; DB 9; Length 610;
Best Local Similarity 95.0%; Pred. No. 1e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
OY 1 TGTACTCAGGGGATGAGAA 20
||||| |||||||||
DB 64 TGTACTTAAAGGATGAGAA 83
RESULT 6
LOCUS BG679991 782 bp mRNA linear EST 01-MAY-2001
DEFINITION 602626709F1 NCI_CGAP_Skn4 Homo sapiens cDNA clone IMAGE:4751499 5',
mRNA sequence.
ACCESSION BG679991
VERSION BG679991.1 GI:13911388
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE NIH-MGC
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs@mail.nih.gov
Tissue Procurement: James Cleaver, M.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNU)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNU at:
http://image.llnl.gov
Plate: ILAM10608 row: a column: 04
High quality sequence start: 6
High quality sequence stop: 781.
FEATURES
Source
Location/Qualifiers
1..782
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4751499"
/clone_lib="NCI CGAP_Skn4"
/tissue_type="squamous cell carcinoma"
/lab_host="DH10B (T1 phage-resistant)"
/note="Organ: skin; Vector: pCMV-SPOK6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.5kb. Library constructed by Life
Technologies. Note: this is a NCI_CGAP Library."

BASE COUNT 179 a 194 c 231 g 178 t
ORIGIN
Query Match 92.0%; Score 18.4; DB 12; Length 782;
Best Local Similarity 95.0%; Pred. No. 1.1e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 TGTACTCAGGGATGAGAA 20
|||||
Db 127 TGTATTCAGGGATGAGAA 146
RESULT 7 1007 bp mRNA linear EST 15-MAY-2001
LOCUS BG749494 602707627f1 NIH_MGC_43 Homo sapiens cDNA IMAGE:484167 5',
DEFINITION mRNA sequence.
ACCESSION BG749494 GI:14060147
VERSION BG749494.1 GI:14060147
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE 1 (bases 1 to 1007)
AUTHORS NIH-MGC <http://mhc.nci.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LNCM1681 row: f column: 08
High quality sequence stop: 714.
Location/Qualifiers
1..1007
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:484167"
/clone_lib="NIH_MGC_43"
/issue_type="normal pigmented retinal epithelium"
/lab_host="DH10B (phage-resistant)"
/note="Organ: eye; Vector: pORF7; Site_1: XhoI; Site_2:
EcoRI; cDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adaptor: GGCACGAG(G). Library constructed by Ling Hong
in the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-CDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies).
Note: this is a NIH_MGC library."

BASE COUNT 232 a 270 c 309 g 196 t
ORIGIN
Query Match 92.0%; Score 18.4; DB 12; Length 1007;
Best Local Similarity 95.0%; Pred. No. 1.2e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 TGTACTCAGGGATGAGAA 20
|||||
Db 140 TGTATTCAGGGATGAGAA 159
RESULT 8 446 bp DNA linear GSS 09-NOV-1999
LOCUS AQ880436 446 bp DNA linear GSS 09-NOV-1999
DEFINITION HS 5044.A2.D10.SP6E.RPCI-11 Human Male BAC Library Homo sapiens
genomic clone Plate-8812 Col-20 Row-G, DNA sequence.
ACCESSION AQ880436

VERSION AQ880436.1 GI:6311903
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE 1 (bases 1 to 446)
AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
Hood,L.
TITLE Sequence-tagged connectors: A sequence approach to mapping and
scanning the human genome
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
MEDLINE 99380589
COMMENT Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Clones are derived from the human BAC library RPCI-11. For BAC
library availability, please contact Pieter de Jong
(pieterdejong.med.bufileo.edu). Clones may be purchased from
BACPAC Resources (http://bacpac.med.bufileo.edu/ordering_bac.htm)
or from Research Genetics (info@resgen.com). BAC end web server:
<http://www.htsc.washington.edu>
Plate: 8812 row: G column: 20
Seq primer: SP6
Class: BAC ends
High quality sequence stop: 446.
Location/Qualifiers
1..446
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Plate-8812 Col-20 Row-G"
/clone_lib="RPCI-11 Human Male BAC Library"
/sex="male"
/note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
Male blood DNA was isolated from one randomly chosen donor
and partially digested with a combination of EcoRI and
EcoRI methylase. Size selected DNA was cloned into the
pBACe3.6 vector at EcoRI sites"

BASE COUNT 116 a 101 c 68 g 159 t
ORIGIN
Query Match 87.0%; Score 17.4; DB 17; Length 446;
Best Local Similarity 94.7%; Pred. No. 2.9e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 2 GTTACTCAGGGATGAGAA 20
|||||
Db 362 GCTACTCAGGGATGAGAA 344
RESULT 9 560 bp DNA linear GSS 24-APR-1999
LOCUS AQ481409 560 bp DNA linear GSS 24-APR-1999
DEFINITION RPCI-11-241H7.TJ RPCI-11 Homo sapiens genomic clone RPCI-11-241H7,
DNA sequence.
ACCESSION AQ481409
VERSION AQ481409.1 GI:4668813
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE 1 (bases 1 to 560)
AUTHORS Zhao,S., Adams,M.D., Nierman,W., Malek,J., de Jong,P. and Venter
J.C.
TITLE Use of BAC End Sequences from Library RPCI-11 for Sequence-Ready
Map Building (1997)
JOURNAL Unpublished (1997)


```

QY      1 TGTACTCAGGGGATGAGAA 20
      .  |||||  |||  |||||
Db     291 TGTACTCATGGGGTGAGAA 277

```

Tue Jul 1 11:47:46 2003

us-10-053-662a-29.rst

Search completed: July 1, 2003, 07:03:41
Job time : 909.615 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 1, 2003, 05:18:35 ; Search time 556.359 Seconds
(without alignments)
993.879 Million cell updates/sec

Title: US-10-053-662A-30

Perfect score: 19

Sequence: 1 ctgggggcagcttattgcac 19

Scoring table: IDENTITY NUC

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : GenBank1.*
1: gb_ba.*
2: gb_hg.*
3: gb_in.*
4: gb_on.*
5: gb_ov.*
6: gb_pat.*
7: gb_ph.*
8: gb_pl.*
9: gb_pr.*
10: gb_ro.*
11: gb_sts.*
12: gb_sy.*
13: gb_un.*
14: gb_vl.*
15: em_ba.*
16: em_fun.*
17: em_hum.*
18: em_in.*
19: em_mu.*
20: em_on.*
21: em_or.*
22: em_ov.*
23: em_pat.*
24: em_ph.*
25: em_pl.*
26: em_ro.*
27: em_sts.*
28: em_un.*
29: em_vl.*
30: em_htg_hum.*
31: em_htg_inv.*
32: em_htg_other.*
33: em_htg_mus.*
34: em_htg_pln.*
35: em_htg_rtd.*
36: em_htg_mam.*
37: em_htg_vrl.*
38: em_sy.*
39: em_htgo_hum.*
40: em_htgo_mus.*
41: em_htgo_other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	17.4	91.6	67605	2	AC096816
C 2	17.4	91.6	100418	2	HS1028E2
C 3	17.4	91.6	102165	2	AC083819
C 4	17.4	91.6	158149	2	AC069000
5	17.4	91.6	163218	9	AL445989
6	17.4	91.6	171012	2	AC016061
7	17.4	91.6	202509	2	AC129875
C 8	17	89.5	168435	2	AC098629
C 9	17	89.5	170238	2	AC117955
C 10	16.4	86.3	2785	2	AK091176
C 11	16.4	86.3	7752	3	AB058743
C 12	16.4	86.3	30102	3	AF000195
C 13	16.4	86.3	38525	3	AF003145
C 14	16.4	86.3	42325	1	U00015
C 15	16.4	86.3	65707	2	AC101474
C 16	16.4	86.3	78576	2	AC127990
C 17	16.4	86.3	93307	2	AC125364
C 18	16.4	86.3	107889	9	AC004519
C 19	16.4	86.3	128484	9	AC025270
C 20	16.4	86.3	160394	2	AC012273
C 21	16.4	86.3	162495	9	AC009996
C 22	16.4	86.3	165423	2	AC120897
C 23	16.4	86.3	166975	9	AC069384
C 24	16.4	86.3	175465	2	AC009821
C 25	16.4	86.3	206644	2	AL732501
C 26	16.4	86.3	344050	1	MEPRTN2
C 27	16	84.2	80398	9	AL513485
C 28	16	84.2	154101	2	AL537565
C 29	16	84.2	163284	9	AL137000
C 30	16	84.2	179024	2	AC112551
C 31	16	84.2	201254	2	AL626767
C 32	16	84.2	341560	2	AL596304
C 33	15.8	83.2	424	9	AF424831
C 34	15.8	83.2	1119	8	AF247165
C 35	15.8	83.2	1138	5	AF370640
C 36	15.8	83.2	1490	6	AX011679
C 37	15.8	83.2	2529	9	AF152487
C 38	15.8	83.2	2592	9	AB033004
C 39	15.8	83.2	2733	8	SCYLR12IC
C 40	15.8	83.2	4020	9	HSALJ7610
C 41	15.8	83.2	5260	9	AF152317
C 42	15.8	83.2	6387	9	AB002343
C 43	15.8	83.2	6868	9	HSM803621
C 44	15.8	83.2	23223	8	YSCLE9233
C 45	15.8	83.2	31694	9	AL353606

ALIGNMENTS

RESULT 1
AC096816/c 67605 bp DNA linear HTG 11-JUL-2002
LOCUS Rattus norvegicus clone CH230-171E18, *** SEQUENCING IN PROGRESS
DEFINITION *** 40 unordered pieces.
ACCESSION AC096816 GI:21723712
VERSION AC096816.3
KEYWORDS HTG, HTGS, PHASE1.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE 1 (bases 1 to 67605)
AUTHORS Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-Osman,F.R., Allen,C.,

Alsbrooks, S.L., Amaralunge, H.C., Are, J.R., Ayele, M., Banks, T.,
 Barbieri, J., Benton, J., Bimage, K., Blankenburg, K., Bonnin, D.,
 Bouche, J., Bowles, S., Brieva, M., Brown, E., Brown, M., Bryant, N.P.,
 Buhay, C., Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C.,
 Carron, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D.,
 Chen, G., Chen, R., Chen, Z., Chowdhry, I., Christopoulos, C.,
 Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R.,
 Davila, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A.,
 Delaney, K.R., Delgado, O., Denn, A.L., Ding, Y., Dinh, K.H.,
 Douthett, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J.,
 Earhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escotto, M.,
 Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P.,
 Gabsi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R.,
 Gorell, J.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K.,
 Harris, C., Harris, K., Hart, M., Haylak, P., Hawes, A., Hernandez, J.,
 Hernandez, O., Hodgson, A., Hognes, M., Holloway, C., Hollins, B.,
 Homsi, F., Howard, S., Huber, J., Hulyk, S., Hume, J., Jackson, L.E.,
 Jacobson, B., Jia, Y., Johnson, R., Jolyet, S., Joudah, S.,
 Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovar, C.,
 Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L.,
 Li, J., Li, Z., Lichtarge, O., Lien, C., Liu, J., Liu, W., Louiseged, H.,
 Lozano, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J.,
 Maheshwari, M., Mapua, P., Martin, R., Marindale, A., Martinez, E.,
 Massey, E., Mawhney, E., McLeod, M.P., Meador, M., Mei, G., Metzker, M.,
 Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, K., Morris, S.,
 Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N.,
 Nguyen, N., Nickerson, E., Nwokenkwo, S., Ogih, M., Okunodu, G.,
 Orangunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L.,
 Peters, L., Pickens, R., Primus, E., Pul, L., Quiles, M., Ren, Y.,
 Rivers, M., Rojas, A., Rojibokan, I., Rolfe, M., Ruiz, S., Saverly, G.,
 Scherer, S., Scott, G., Shen, H., Shoshitari, N., Sisson, I.,
 Sodergren, E., Sonaite, T., Sparks, A., Stanley, H., Stone, H.,
 Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H.,
 Tansy, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S.,
 Usmari, K., Vasquez, L., Vera, V., Villalón, D., Vinson, R., Wang, O.,
 Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S.,
 Williams, G., Williamson, A., Wleczek, R., Wooden, S., Worley, K.,
 Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D.,
 Weinstock, G., and Gibbs, R.

Unpublished
 Direct Submission
 2 (bases 1 to 67605)
 Worley, K.C.

Direct Submission
 Submitted (28-SEP-2001) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 3 (bases 1 to 67605)
 Worley, K.C.

Direct Submission
 Submitted (11-JUL-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 On Jul 10, 2002 this sequence version replaced gi:17944640.

----- Genome Center

Center: Baylor College of Medicine
 Center code: BCM
 Web site: <http://www.hgsc.bcm.tmc.edu/>
 Contact: hgsc-help@bcm.tmc.edu

----- Project Information

Center project name: GRVY
 Center clone name: CH230-171E18

----- Summary Statistics

Sequencing vector: Plasmid;
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.990329
 Consensus quality: 36667 bases at least Q40
 Consensus quality: 30535 bases at least Q30
 Consensus quality: 33391 bases at least Q20

* NOTE: Estimated insert size may differ from sequence length
 * (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
 * NOTE: This is a 'working draft' sequence. It currently

* consists of 40 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1
 1006: contig of 1006 bp in length
 1107: gap of unknown length
 1107: contig of 1484 bp in length
 2590: gap of unknown length
 2591: gap of unknown length
 2591: contig of 1251 bp in length
 3941: gap of unknown length
 4041: gap of unknown length
 5216: contig of 1175 bp in length
 5217: gap of unknown length
 6779: contig of 1463 bp in length
 6780: gap of unknown length
 7907: contig of 1028 bp in length
 7908: gap of unknown length
 8007: gap of unknown length
 8008: contig of 1017 bp in length
 9024: gap of unknown length
 9124: gap of unknown length
 10466: contig of 1342 bp in length
 10467: gap of unknown length
 10567: gap of unknown length
 11821: contig of 1255 bp in length
 11822: gap of unknown length
 11921: gap of unknown length
 13244: contig of 1323 bp in length
 13345: gap of unknown length
 15223: contig of 1879 bp in length
 15324: gap of unknown length
 15324: contig of 1500 bp in length
 16824: gap of unknown length
 16824: contig of 1078 bp in length
 18001: gap of unknown length
 18101: gap of unknown length
 19379: contig of 1278 bp in length
 19479: gap of unknown length
 21372: contig of 1893 bp in length
 21472: gap of unknown length
 22886: contig of 1414 bp in length
 22887: gap of unknown length
 22986: gap of unknown length
 24569: contig of 1583 bp in length
 24570: gap of unknown length
 26748: contig of 2079 bp in length
 26749: gap of unknown length
 26848: gap of unknown length
 28102: contig of 1254 bp in length
 28202: gap of unknown length
 29707: contig of 1505 bp in length
 29807: gap of unknown length
 30900: contig of 1093 bp in length
 30901: gap of unknown length
 31001: gap of unknown length
 32406: contig of 1406 bp in length
 32407: gap of unknown length
 32506: gap of unknown length
 34962: contig of 2456 bp in length
 35062: gap of unknown length
 35063: contig of 1777 bp in length
 36839: gap of unknown length
 36840: gap of unknown length
 38727: contig of 1788 bp in length
 38827: gap of unknown length
 38828: gap of unknown length
 40324: contig of 1496 bp in length
 40324: gap of unknown length
 41852: contig of 1429 bp in length
 41853: gap of unknown length
 41952: gap of unknown length
 43772: contig of 1820 bp in length
 43773: gap of unknown length
 43872: gap of unknown length
 45456: contig of 1584 bp in length
 45457: gap of unknown length
 45556: gap of unknown length
 47553: contig of 1997 bp in length
 47554: gap of unknown length
 47653: gap of unknown length
 48905: contig of 1252 bp in length
 48906: gap of unknown length
 49005: gap of unknown length
 50575: contig of 1570 bp in length
 50675: gap of unknown length
 51897: contig of 1222 bp in length
 51897: gap of unknown length

* 51998 54197: contig of 2200 bp in length
* 54198 54297: gap of unknown length
* 54298 55860: contig of 1563 bp in length
* 55861 55960: gap of unknown length
* 55961 58193: contig of 2233 bp in length
* 58194 58293: gap of unknown length
* 58294 58961: contig of 1568 bp in length
* 58962 59961: gap of unknown length
* 59962 62224: contig of 2263 bp in length
* 62225 62324: gap of unknown length
* 62325 64813: contig of 2489 bp in length
* 64814 67605: gap of unknown length
* 64914 67605: contig of 2692 bp in length.
Location/Qualifiers
1. 67605
/organism="Rattus norvegicus"
/db_xref="taxon:10116"
/clone="CH230-171E18"
BASE COUNT 15603 a 14128 c 14795 g 14669 t 8410 others
ORIGIN
Query Match 91.6%; Score 17.4; DB 2; Length 67605;
Best Local Similarity 94.7%; Pred. No. 70;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
OY 1 CTGGGGGAGTATTATGAC 19
|||||
Db 24769 CTGGGGGAGTATTATGAC 24751
RESULT 2
HS1026E2 100418 bp DNA linear PRI 23-NOV-1999
LOCUS HS1026E2/c
DEFINITION Human DNA sequence from clone 1026E2 on chromosome 1q24.1-25.3 EST,
CA repeat, STS, GSS, complete sequence.
ACCESSION AL022143
VERSION AL022143.1 GI:3319672
KEYWORDS HTG; repeat polymorphism.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 100418)
Wilkinson, J.
Direct Submission
Submitted (08-JUL-1998) E-mail enquiries: humquerry@sanger.ac.uk
Clone requests: clonerequest@sanger.ac.uk
On Jul 15, 1998 this sequence version replaced gi:2969924.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences
with only a small overlap as described above. IMPORTANT: This
sequence is not the entire insert of clone 1026E2. It may be
shorter because we only sequence overlapping sections once, or
longer because we arrange for a small overlap between neighbouring
submissions.
The true left end of clone 964D12 is at 100315 in this sequence.
The true right end of clone 35C21 is at 59922 in this sequence.
This sequence has been finished according to sequence map criteria
as follows. An attempt is made to resolve all sequencing problems,
such as compressions and repeats, but not necessarily within known
annotated human repeat sequence elements (e.g. Alu). Where the
sequence is ambiguous, there is an annotation using the 'unsure'
feature key.
This sequence was generated from part of bacterial clone contigs of
human chromosome 1, constructed by the Sanger Centre Chromosome 1
Mapping Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Ch1
1026E2 is from the library RPI5 constructed at the Roswell Park
Cancer Institute by the group of Pieter de Jong. For further
details see http://bacpac.med.buffalo.edu/VECTOR: pcypac2.

FEATURES
source
Location/Qualifiers
1. 100418
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="1"
/map="q24.1-25.3"
/clone="RP5-1026E2"
/clone_1lb="RPI5-5"
9. 1897
/note="Retroviral Sequence"
2018. 2177
/note="2 copies 80 mer 95% conserved"
2052. >2289
/note="match: EST AA601759 clone IMAGE:1099629"
2176. 2339
/note="2 copies 82 mer 85% conserved"
prim_transcript
/note="match: 3' EST AA783035 clone 1375023"
repeat_region
/note="match: 320. 3367)
/note="PR5 repeat: matches 288. 142 of consensus"
complement(325. 3898)
/note="PR5 repeat: matches 2427. 1800 of consensus"
complement(3900. 3945)
/note="PR5 repeat: matches 1397. 1353 of consensus"
3987. 4286
/note="match: 253809 STS containing (CA) repeat"
4102. 4127
/note="13 copies of AC 100% conserved; differs from
253809"
5110. 5398
/note="AluB repeat: matches 12. 302 of consensus"
complement(5436. 5576)
/note="MR2 repeat: matches 146. 4 of consensus"
5606. 5937
/note="MR2 repeat: matches 8. 344 of consensus"
complement(6655. 7102)
/note="match: GSS B82054"
complement(7192. 7423)
/note="11PA10 repeat: matches 900. 662 of consensus"
9003. 9267
/note="MR repeat: matches 2. 262 of consensus"
complement(13451. 13902)
/note="MLTID repeat: matches 505. 44 of consensus"
14327. 14444
/note="MR repeat: matches 47. 164 of consensus"
complement(14857. 14933)
/note="MR5A repeat: matches 179. 37 of consensus"
15170. 15209
/note="10 copies 4 mer cat 83% conserved"
complement(19889. 20167)
/note="MR33 repeat: matches 302. 26 of consensus"
24403. 24428
/note="13 copies 2 mer tt 92% conserved"
24433. 24855
/note="MTF2_internal repeat: matches 1739. 2174 of
consensus"
complement(24871. 25777)
/note="L1PA15 repeat: matches 904. 1 of consensus"
complement(25626. 26986)
/note="L1 repeat: matches 5390. 3989 of consensus"
26985. 27537
/note="L1 repeat: matches 2116. 2707 of consensus"
complement(27553. 27820)
/note="Alu repeat: matches 301. 34 of consensus"
27822. 29084
/note="L1 repeat: matches 2700. 3980 of consensus"
29060. 29679
/note="MT2_internal repeat: matches 2145. 2774 of
consensus"
29763. 30148
/note="STD repeat: matches 3. 394 of consensus"
30148. 30776
/note="MT2_internal repeat: matches 2859. 3495 of

```

repeat_region      consensus
30922..31355
/note="MIR2 internal repeat: matches 3710..4142 of
consensus"
repeat_region      /note="L1 repeat: matches 4149..5390 of consensus"
32755..33606
/note="L1MB6 repeat: matches 1..912 of consensus"
repeat_region      /note="MIR repeat: matches 247..32 of consensus"
35440..35544
/note="MIR repeat: matches 4..109 of consensus"
36090..36149
/note="MIR repeat: matches 146..48 of consensus"
367067..37716
/note="MIR repeat: matches 297..2 of consensus"
38661..40016
/note="MIR repeat: matches 300..1 of consensus"
40911..42821
/note="MIR repeat: matches 157..48 of consensus"
41587..41620
/note="MIR repeat: matches 143..108 of consensus"
41587..41620
/note="MIR repeat: matches 9..128 of consensus"
41587..41620
/note="MIR repeat: matches 206..248 of consensus"
41587..41620
/note="MIR repeat: matches 192..52 of consensus"
41587..41620
/note="MIR repeat: matches 150..81 of consensus"
41587..41620
/note="MIR repeat: matches 146..18 of consensus"
41587..41620
/note="MIR repeat: matches 301..1 of consensus"
41587..41620
/note="MIR repeat: matches 300..1 of consensus"
41587..41620
/note="MIR repeat: matches 187..66 of consensus"
41587..41620
/note="MIR repeat: matches 5047..3929 of consensus"
41587..41620
/note="MIR repeat: matches 293..1 of consensus"
41587..41620
/note="MIR repeat: matches 2418..1 of consensus"
41587..41620

```

```

repeat_region      /note="MIR2 repeat: matches 15..146 of consensus"
65656..65790
/note="MIR2 repeat: matches 4..144 of consensus"
65829..66276
/note="match: GSS A0002124"
repeat_region      67773..67839
/note="MIR repeat: matches 79..146 of consensus"
68924..68987
/note="32 copies 2 mer gt 83% conserved"
69506..70033
/note="L1 repeat: matches 4843..5389 of consensus"
69893..70891
/note="L1MB6 repeat: matches 1..1047 of consensus"

Query Match      91.6%; Score 17.4; DB 9; Length 100418;
Best Local Similarity 94.7%; Pred. No. 69;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY      1 CTGGGGGCGATTATTCAC 19
Db      90926 CTGGGCTAGTATTCAC 90908

RESULT 3
AC083819/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
Metzger,M.L., Lewis,L.R., Hume,J., Edwards,C., Harris,C.,
Dederich,D., Thomas,S., Okunou,G., Carllock,C., Garner,T.,
Addison,S., Pace,A., Williams,G., Bonin,D., Brooks,A., Brown,J.,
Buhay,C., Bunac,C., Burkett,C., Chacko,J., Chen,G., Chen,Z.,
Cox,C., Davis,C., Delgado,O., Ding,Y., Dugan-Rocha,S.,
Fernandez,C., Ferraguto,D., Fortum-Tansey,J., Gill,R.,
Gorell,J.H., Gunaratne,P., Haller,G., Hernandez,J., Hognes,M.,
Hosak,H., Hou,X., Huber,J., Jackson,L., Jia,Y., Kelly,J., Kelly,S.,
Kovay,C., Liu,J., Liu,W., Loulseged,H., Lozano,R.J., Martin,R.,
Massey,E., McLeod,M.P., Mel,G., Moore,S., Morgan,M., Morris,S.,
Neal,D., Nelson,A., Nguyen,R., Nguyen,N., Oquh,M., Parish,B.,
Perez,L., Reiter,D., Say,J., Shen,H., Vasquez,L., Watlington,S.,
Williamson,A., Wrensford,G., Zhou,X., Bouck,J., Hodgson,A.,
Muzny,D.M., Rives,M., Scherer,S., Sodergren,E., Weinstock,G.,
Worley,K. and Gibbs,R.
Direct Submission
Unpublished
2 (bases 1 to 102165)
Worley,K.C.
Direct Submission
Submitted (01-OCT-2000) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Oct 16, 2000 this sequence version replaced gi:10801956.
----- Genome Center -----
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information -----
Center project name: MARO
Center clone name: RP23-285F20
----- Summary Statistics -----
Sequencing vector: M13; 108821
Chemistry: Dye-primer Bodipy; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 95179 bases at least Q40

```



```

repeat_region 28192..28338
/Note="MER3 repeat: matches 10..151 of consensus"
repeat_region 30716..31073
/Note="179 copies 2 mer tt 54% conserved"
repeat_region 32832..33392
/Note="LTR2CA repeat: matches 1..508 of consensus"
repeat_region 33643..33753
/Note="LTR16C repeat: matches 262..378 of consensus"
repeat_region 34208..34245
/Note="19 copies 2 mer aa 78% conserved"
repeat_region 34492..34684
/Note="HERVL repeat: matches 3334..3634 of consensus"
repeat_region 42139..42202
/Note="MIR repeat: matches 79..142 of consensus"
repeat_region 42232..42383
/Note="76 copies 2 mer ta 63% conserved"
repeat_region 43902..43926
/Note="L1 repeat: matches 4804..4829 of consensus"
repeat_region 43927..44474
/Note="L1MA8 repeat: matches 5733..6287 of consensus"
repeat_region 44475..44631
/Note="L1 repeat: matches 4658..4804 of consensus"
repeat_region 45347..45650
/Note="Alu10 repeat: matches 4..305 of consensus"
repeat_region 46847..46940
/Note="47 copies 2 mer tt 62% conserved"
repeat_region 47049..47110
/Note="HAL1 repeat: matches 706..761 of consensus"
repeat_region 48421..48470
/Note="25 copies 2 mer ac 90% conserved"
repeat_region 49812..50212
/Note="LTR1A1 repeat: matches 1..365 of consensus"
repeat_region 52787..52850
/Note="32 copies 2 mer ta 76% conserved"
repeat_region 53040..53341
/Note="Alu1 repeat: matches 1..303 of consensus"
repeat_region 54732..55698
/Note="L1PA8 repeat: matches 4155..6158 of consensus"
repeat_region 56696..57975
/Note="L1PA8 repeat: matches 2725..3982 of consensus"
repeat_region 57976..58471
/Note="L2 repeat: matches 2268..2748 of consensus"
repeat_region 60694..60783
/Note="L2 repeat: matches 2317..2407 of consensus"
repeat_region 61514..61934
/Note="MER31A repeat: matches 3..429 of consensus"
repeat_region 61944..62082
/Note="L2 repeat: matches 1770..1915 of consensus"
repeat_region 62161..62453
/Note="Alu10 repeat: matches 5..293 of consensus"
repeat_region 62568..62963
/Note="L1M4 repeat: matches 2131..2549 of consensus"
repeat_region 63150..63533
/Note="L1M4 repeat: matches 1516..1906 of consensus"
repeat_region 63638..63689
/Note="MER47 repeat: matches 143..194 of consensus"
repeat_region 63687..63744
/Note="MER47 repeat: matches 2244..2303 of consensus"
repeat_region 63775..63981
/Note="L1M4 repeat: matches 1278..1497 of consensus"
repeat_region 63986..64486
/Note="L1M4 repeat: matches 5838..6330 of consensus"
repeat_region 64487..64834
/Note="LTR1A1 repeat: matches 1..365 of consensus"
repeat_region 64835..65482
/Note="L1M2 repeat: matches 5164..5838 of consensus"
repeat_region 65489..65820
/Note="L1M8 repeat: matches 5810..6141 of consensus"
repeat_region 65883..66146
/Note="L1M2 repeat: matches 973..1251 of consensus"
repeat_region 66884..67288
/Note="WSTB repeat: matches 1..426 of consensus"
repeat_region 67573..67853

```

```

repeat_region 69319..69598
/Note="LTR1E repeat: matches 54..364 of consensus"
repeat_region 71293..71849
/Note="L1MA9 repeat: matches 4999..5555 of consensus"
repeat_region 71850..72551
/Note="L1MA9 repeat: matches 5608..6308 of consensus"
repeat_region 75195..75277
/Note="L2 repeat: matches 2660..2750 of consensus"
repeat_region 76986..77091
/Note="L1PA3 repeat: matches 6041..6146 of consensus"
repeat_region 77667..77864
/Note="99 copies 2 mer aa 56% conserved"
repeat_region 78135..80473
/Note="TIGER1 repeat: matches 1..2407 of consensus"

Query Match 91.6%; Score 17.4; DB 9; Length 163218;
Best Local Similarity 94.7%; Pred. No. 69;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 CTGGGCGCAGTATTGCAC 19
DB 8349 CTGGGCGCAGTATTGCC 8367

RESULT 6
LOCUS AC016061
DEFINITION Homo sapiens clone RP11-6N9, WORKING DRAFT SEQUENCE, 15 unordered
pieces:
ACCESSION AC016061.3 GI:7239536
VERSION AC016061.3
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
1 (bases 1 to 171012)
Birten,B., Linton,L., Nusbaum,C. and Lander,E.
Homo sapiens chromosome, clone RP11-6N9
Unpublished
2 (bases 1 to 171012)
Birten,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
Baldwin,J., Barns,N., Beckert,R., Boguslavsky,L., Boukhalter,B.,
Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,
Cooke,P., Deatellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
Ferreira,P., Fitzhugh,W., Forrest,C., Funke,R., Gage,D.,
Gallagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
Lehoczky,J., Liu,C., Locke,K., Macdonald,P., Marquis,N.,
McKean,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrum,J.,
Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
Stange-Thomann,N., Stojanovic,N., Stojanovic,A., Talamas,J.,
Testafaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
Direct Submission
Submitted (20-NOV-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Mar 14, 2000 this sequence version replaced g1:6705863.
All repeats were identified using RepeatMasker:
Smit,A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www.seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L2912
Center clone name: 6_N_9
----- Summary Statistics
Sequencing vector: M13; M77815; 100% of reads

```

Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 161177 bases at least Q40
Consensus quality: 166152 bases at least Q30
Consensus quality: 168156 bases at least Q20
Insert size: 151000; agarose-fp
Insert size: 169612; sum-of-contrigs
Quality coverage: 4.9 in Q20 bases; agarose-fp
Quality coverage: 4.3 in Q20 bases; sum-of-contrigs

NOTE: This is a 'working draft' sequence. It currently
* consists of 15 contrigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contrigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 1250: contrig of 1250 bp in length
1351 1350: gap of 100 bp
1351 2445: contrig of 1095 bp in length
2446 2545: gap of 100 bp
2546 3546: contrig of 1001 bp in length
3547 3646: gap of 100 bp
3647 5950: contrig of 2304 bp in length
5951 6050: gap of 100 bp
6051 9159: contrig of 3109 bp in length
9160 9259: gap of 100 bp
9260 11925: contrig of 2666 bp in length
11926 12025: gap of 100 bp
12026 16254: contrig of 4229 bp in length
16255 16354: gap of 100 bp
16355 21948: contrig of 5594 bp in length
21949 22048: gap of 100 bp
22049 29743: contrig of 7695 bp in length
29744 29843: gap of 100 bp
29844 38782: contrig of 8939 bp in length
38783 38882: gap of 100 bp
38883 53976: contrig of 15094 bp in length
53977 54076: gap of 100 bp
54077 71196: contrig of 17120 bp in length
71197 71296: gap of 100 bp
71297 101039: contrig of 29743 bp in length
101040 101139: gap of 100 bp
101140 129583: contrig of 28444 bp in length
129584 129683: gap of 100 bp
129684 171012: contrig of 41329 bp in length.

FEATURES

source
1.171012
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="RPC1-11 Human Male BAC"
1.1250
/note="assembly_fragment"
misc_feature
1351..2445
/note="assembly_fragment"
misc_feature
2546..3546
/note="assembly_fragment"
vector_end:sp6
vector_side:right
3647..5950
/note="assembly_fragment"
misc_feature
6051..9159
/note="assembly_fragment"
misc_feature
9260..11925
/note="assembly_fragment"
clone_end:t7
vector_side:right
12026..16254
/note="assembly_fragment"
misc_feature
16355..21948
/note="assembly_fragment"

misc_feature 22049..29743
/note="assembly_fragment"
misc_feature 29844..38782
/note="assembly_fragment"
misc_feature 38883..53976
/note="assembly_fragment"
misc_feature 54077..71196
/note="assembly_fragment"
misc_feature 71297..101039
/note="assembly_fragment"
misc_feature 101140..129583
/note="assembly_fragment"
misc_feature 129684..171012
/note="assembly_fragment"

BASE COUNT 55604 a 29837 c 29224 g 54946 t 1401 others

Query Match 91.6%; Score 17.4; DB 2; Length 171012;
Best Local Similarity 94.7%; Pred. No. 67;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 21022 CTGGGGCAGTATTTCAC 19
1 CTGGGGCAGTATTTCAC 19
|||||||

RESULT 7
AC129875/c 202509 bp DNA linear HTG 04-AUG-2002

LOCUS Rattus norvegicus clone CH230-11363, *** SEQUENCING IN PROGRESS

DEFINITION *** 65 unordered pieces.

ACCESSION AC129875.1 GI:22095101

VERSION AC129875.1

KEYWORDS HTGS: PHASE1.

SOURCE Norway rat.

ORGANISM Rattus norvegicus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;

Rattus.

1 (bases 1 to 202509)

REFERENCE

AUTHORS Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,

Alshrocks,S.L., Amartunge,H.C., Are,J.R., Ayala,M., Banks,T.,

Barbieri,J., Benton,J., Bimaye,K., Blankenburg,K., Bonnin,D.,

Bouck,U., Bowle,S., Brileva,M., Brown,E., Brown,M., Bryant,N.P.,

Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,

Carroll,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,

Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,

Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,

Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,

Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,

Douthwalte,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,

Edwards,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,

Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frazz,P.,

Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,

Correll,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,

Harris,C., Harris,K., Hart,M., Haylak,P., Hawes,A., Hernandez,J.,

Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,

Homsel,F., Howard,S., Huber,J., Huiyk,S., Hume,J., Jackson,L.E.,

Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,

Karlsson,E., Kelly,S., Khan,U., King,L., Korvab,J., Kovar,C.,

Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,

Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Louisedge,H.,

Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,

Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,

Massey,E., Mawhney,E., Mcleod,M.P., Meador,M., Mel,G., Metzger,M.,

Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,

Moser,M., Neal,D., Newton,J., Newton,S., Ogun,M., Okunolu,G.,

Oragunye,N., Oviado,R., Pace,A., Payton,B., Peery,J., Perez,L.,

Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,

Rives,M., Rojas,A., Rojibokan,I., Rolife,M., Ruiz,S., Savery,G.,

Scherer,S., Scott,G., Shen,H., Shoostari,N., Sisson,I.,

Sodergren,E., Sonalke,T., Sparks,A., Stanley,H., Stone,H.,

*	122550	gap of unknown length
*	122550	gap of unknown length
*	122550	contig of 4933 bp in length
*	127583	gap of unknown length
*	127583	contig of 3812 bp in length
*	131495	gap of unknown length
*	131495	contig of 4258 bp in length
*	135752	gap of unknown length
*	135752	gap of unknown length
*	141135	contig of 5263 bp in length
*	141135	gap of unknown length

Query Match	91.6%;	Score 17.4;	DB 2;	length 202509;
Best Local Similarity	94.7%;	Pred. No. 67;		
Matches 18; Conservative	0;	Mismatches 1;	Indels 0;	Gaps 0;

QY	1	CTGGGGGCGATTATTCAC	19
Db	151028	CTGGGGTCAGTATTGCAC	1510100

RESULT 8					
AC098629/c					
LOCUS	AC098629	168435 bp	DNA	linear	HTG 23-JUL-2002
DEFINITION	Rattus norvegicus clone CH230-108I22, *** SEQUENCING IN PROGRESS				
	***, 62 unordered pieces.				
REVISION	NO				
DESCRIPTION	NO				

REFERENCE
AUTHORS

1 (bases 1 to 168435)

Munzy, D.M., Adams, C., Adio-Oduola, B., Alt-oman, F.R., Allen, C., Albrooks, S.L., Amaratunge, H.C., Are, J.R., Ayele, M., Banks, T., Barbieri, J., Benson, J., Blinage, K., Blankenburg, K., Bonini, D., Borch, J., Bowie, S., Brivela, M., Brown, M., Brown, M., Bryant, N. P., Bunyah, C., Burich, P., Burkett, C., Butrell, K.L., Byrd, N.C., Caron, T. F., Carter, M., Cavazos, S. R., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Z., Chowdhury, I., Christopoulos, C., Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R., Davila, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A., Delaney, K.R., Delgado, O., Denn, A.L., Ding, Y., Din, H.H., Douthwaite, K.J., Draper, H., Dugan-Rocha, S., Dumbin, K.J., Eamhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escotto, M., Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P., Garris, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R., Gorrell, J. H., Guevara, M., Gunarrine, P., Hale, S., Hamilton, K., Harris, C., Harris, K., Hart, M., Haylak, P., Hawes, A., Hernandez, J., Hernandez, O., Hodgson, A., Hogues, M., Holloway, C., Hollins, B., Homsi, F., Howard, S., Huber, J., Huliyk, S., Hume, J., Jackson, L.E., Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudah, S., Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovar, C., Kratoch, J., Kureshi, A., Landty, N., Deal, B., Lewis, L.C., Lewis, L., Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, Y., Louisseg, H., Lozado, R. J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J., Mareshwari, M., Mapua, P., Martin, R., Martindale, A., Mattheiz, E., Massey, E., Mahoney, E., Mcleod, M. P., Meador, M., Mei, G., Metzger, M., Miner, G., Miner, Z., Mitchell, J., Mohabbat, K., Morgan, M., Morris, S., Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nwokenkwo, S., Ogun, M., Okwundu, G., Oregunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L., Peters, L., Pickens, R., Primus, E., Pu, L. L., Quiles, M., Ren, Y., Rives, M., Rojas, A., Rojibokan, I., Rolfe, M., Ruiz, S., Savery, G., Scherer, S., Scott, G., Shen, H., Shooshitari, N., Sisson, I., Sodergren, E., Sonalke, T., Sparks, A., Stanley, H., Stone, H., Sutton, A., Swalek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H., Tansley, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S., Usmani, K., Vasquez, L., Vera, V., Villalón, D., Vison, R., Wang, Q., Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S., Williams, G., Williamson, A., Wleciyk, R., Woden, S., Worley, K., Wu, C., Wu, Y., Wu, Y. F., Zhou, Y., Zorrilla, S., Nelson, D., Weinstein, G., and Gibbs, R.

TITLE	Direct Submission
JOURNAL	Unpublished
REFERENCE	2 (bases 1 to 168435)
AUTHORS	Worley,K.C.
TITLE	Direct Submission
JOURNAL	Submitted (27-OCT-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
REFERENCE	3 (bases 1 to 168435)
AUTHORS	Worley,K.C.
TITLE	Direct Submission
JOURNAL	Submitted (23-JUL-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
COMMENT	On May 29, 2002 this sequence version replaced gi:17973473.

NOTE: Estimated insert size may differ from sequence length
(see http://www.bhsc.bcm.tmc.edu/docs/genbank_draft_data.html)

NOTE: This is a 'working draft' sequence. It currently
consists of 62 contigs. The true order of the pieces is
is not known and their order in this sequence record is
arbitrary. Gaps between the contigs are represented as
runs of N, but the exact sizes of the gaps are unknown.
This record will be updated with the finished sequence
as soon as it is available and the accession number will
be preserved.

1	1058: contig of 1058 bp in length
1059	1158: gap of unknown length
1159	2246: contig of 1088 bp in length
2247	2346: gap of unknown length
2347	3914: contig of 1568 bp in length
3915	4014: gap of unknown length
4015	5410: contig of 1396 bp in length
5411	5510: gap of unknown length
5511	7015: contig of 1505 bp in length
7016	7115: gap of unknown length
7116	8297: contig of 1182 bp in length
8298	8397: gap of unknown length
8398	9457: contig of 1060 bp in length
9458	9557: gap of unknown length
9558	11326: contig of 1769 bp in length
11327	11426: gap of unknown length
11427	12468: contig of 1042 bp in length
12469	12568: gap of unknown length
12569	13856: contig of 1288 bp in length
13857	13956: gap of unknown length
13957	15471: contig of 1515 bp in length
15472	15571: gap of unknown length
15572	16585: contig of 1014 bp in length
16586	16685: gap of unknown length
16686	17938: contig of 1253 bp in length
17939	18038: gap of unknown length
18039	19446: contig of 1408 bp in length
19447	19546: gap of unknown length
19547	20676: contig of 1130 bp in length
20677	20776: gap of unknown length
20777	22119: contig of 1343 bp in length
22120	22219: gap of unknown length

```

* 22220 23410: contig of 1191 bp in length
* 23411 23510: gap of unknown length
* 23511 24594: contig of 1084 bp in length
* 24595 24694: gap of unknown length
* 24695 25900: contig of 1206 bp in length
* 25901 26000: gap of unknown length
* 26001 27046: contig of 1046 bp in length
* 27047 27146: gap of unknown length
* 27147 28430: contig of 1284 bp in length
* 28431 28530: gap of unknown length
* 28531 29780: contig of 1250 bp in length
* 29781 29880: gap of unknown length
* 29881 31611: contig of 1731 bp in length
* 31612 31712: gap of unknown length
* 31713 33655: contig of 1944 bp in length
* 33656 33755: gap of unknown length
* 33756 35174: contig of 1419 bp in length
* 35175 35274: gap of unknown length
* 35275 37331: contig of 2057 bp in length
* 37332 37431: gap of unknown length
* 37432 39444: contig of 2013 bp in length
* 39445 39544: gap of unknown length
* 39545 40865: contig of 1321 bp in length
* 40866 40965: gap of unknown length
* 40966 43515: contig of 2550 bp in length
* 43516 43615: gap of unknown length
* 43616 45690: contig of 2075 bp in length
* 45691 45791: gap of unknown length
* 45791 46803: contig of 1013 bp in length
* 46804 46903: gap of unknown length
* 46904 49441: contig of 2538 bp in length
* 49442 49541: gap of unknown length
* 49542 51912: contig of 2371 bp in length
* 51913 52012: gap of unknown length
* 52013 54896: contig of 2884 bp in length
* 54897 54996: gap of unknown length
* 54997 56315: contig of 1319 bp in length
* 56316 59301: gap of unknown length
* 59302 59401: gap of unknown length
* 59402 62164: contig of 2763 bp in length
* 62165 62264: gap of unknown length
* 62265 64647: contig of 2383 bp in length
* 64648 64747: gap of unknown length
* 64748 66818: contig of 2071 bp in length
* 66819 66918: gap of unknown length
* 66919 69765: contig of 2847 bp in length
* 69766 69865: gap of unknown length
* 69866 73674: contig of 3809 bp in length
* 73675 73774: gap of unknown length
* 73775 76147: contig of 2373 bp in length
* 76148 76247: gap of unknown length
* 76248 79383: contig of 3136 bp in length
* 79384 79483: gap of unknown length
* 79484 82525: contig of 3042 bp in length
* 82526 82625: gap of unknown length
* 82626 86477: contig of 3852 bp in length
* 86478 86577: gap of unknown length
* 86578 88928: contig of 2351 bp in length
* 88929 89028: gap of unknown length
* 89029 92657: contig of 3629 bp in length
* 92658 92757: gap of unknown length
* 92758 95689: contig of 2932 bp in length
* 95690 95789: gap of unknown length
* 95790 100042: contig of 4253 bp in length
* 100043 100142: gap of unknown length
* 100143 103701: contig of 3559 bp in length
* 103702 103801: gap of unknown length
* 103802 107288: contig of 3487 bp in length
* 107289 107388: gap of unknown length
* 107389 112161: contig of 4773 bp in length
* 112162 112261: gap of unknown length
* 112262 115016: contig of 2755 bp in length

```

```

* 115017 115116: gap of unknown length
* 115117 120847: contig of 5731 bp in length
Query Match 89.58; Score 17; DB 2; Length 168435;
Best Local Similarly 100.0%; Pred. No. 1.1e-02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2 TGGGGGCGATTATGCA 18
Db 156771 TGGGGGCGATTATGCA 156755

```

```

RESULT 9
AC117955/c
LOCUS
DEFINITION
Rattus norvegicus clone CH230-297H21, *** SEQUENCING IN PROGRESS
ACCESSION
AC117955
VERSION
AC117955.4 GI:21745748
KEYWORDS
HTG: HTGS PHASE1.
SOURCE
Norway rat.
ORGANISM
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

```

```

REFERENCE
1 (bases 1 to 170238)
AUTHORS
Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,

```

```

Aisbrooks,S.L., Amaralunge,H.C., Are,J.R., Ayale,M., Banks,T.,
Barbora,J., Benton,J., Bimaga,K., Blankenburg,K., Bonnin,D.,
Bouck,J., Bowie,S., Brileva,M., Brown,E., Brown,M., Bryant,N.P.,
Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
Carion,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinu,H.H.,
Doutwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
Eamhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
Falls,T., Ferraguto,D., Flagg,N., Ford,J.J., Foster,P., Frantz,P.,
Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
Gottrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
Harris,C., Harris,K., Hart,M., Havlak,P., Hayes,A., Hernandez,J.,
Hernandez,O., Hodgson,A., Hognes,M., Holloway,C., Hollins,B.,
Homsli,F., Howard,S., Huber,J., Hulik,S., Hume,J., Jackson,L.E.,
Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
Karlsom,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.,
Li,J., Li,Z., Licharge,O., Lieu,C., Liu,J., Liu,W., Louisedge,H.,
Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
Maheshwari,M., Mapa,P., Martin,R., Martindale,A., Martinez,E.,
Massey,E., Mawhinney,E., McLeod,M.P., Meador,M., Mel,G., Metzker,M.,
Miner,G., Miner,Z., Mitchell,T., Mohabac,K., Morgan,M., Morris,S.,
Mosier,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N.,
Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,L., Perez,L.,
Peters,L., Pickens,R., Plims,E., Pu,L.L., Quiles,M., Ren,Y.,
Rives,M., Rojas,A., Rojibokan,I., Rolfe,M., Ruiz,S., Savery,G.,
Scherer,S., Scott,G., Shen,H., Shooshari,N., Sisson,I.,
Sodergren,E., Sonalke,T., Sparks,A., Stanley,H., Stone,H.,
Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
Tatney,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,
Usmani,K., Vasquez,L., Vera,V., Villalón,D., Vinson,R., Wang,O.,
Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
Williams,G., Williamson,A., Wleczky,R., Wooden,S., Worley,K.,
Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorilla,S., Nelson,D.,
Wainstock,G., and Gibbs,R.

```

```

TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
Unpublished
Direct Submission
2 (bases 1 to 170238)
Submitted (12-Apr-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One

```

REFERENCE
AUTHORS
TITLE
JOURNAL

COMMENT

Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 170238)
Morley, K.C.
Direct Submission
Submitted (18-JUL-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Jul 14, 2002 this sequence version replaced gi:20258124.

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: hgsc-help@bcm.tmc.edu

Project Information

Center project name: GTUN

Center clone name: CH230-297H21

Summary Statistics

Sequencing vector: plasmid:

Chemistry: Dye-terminator Big Dye: 100% of reads

Assembly program: Phrap; version 0.990329

Consensus quality: 129980 bases at least Q40

Consensus quality: 135538 bases at least Q30

Consensus quality: 139850 bases at least Q20

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 43 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence.
* as soon as it is available and the accession number will
* be preserved.

1 1294: contig of 1294 bp in length
* 1295 1394: gap of unknown length
* 1395 2602: contig of 1208 bp in length
* 2603 2702: gap of unknown length
* 2703 4328: contig of 1626 bp in length
* 4328 4429: gap of unknown length
* 4429 6030: contig of 1602 bp in length
* 6030 6130: gap of unknown length
* 6130 7521: contig of 1391 bp in length
* 7521 7621: gap of unknown length
* 7621 7622 8853: contig of 1232 bp in length
* 7622 8854 10205: contig of 1252 bp in length
* 8854 10205 10305: gap of unknown length
* 10205 10305 11476: contig of 1171 bp in length
* 11476 11576: gap of unknown length
* 11576 13046: contig of 1470 bp in length
* 13046 13146: gap of unknown length
* 13146 14641: contig of 1495 bp in length
* 14641 14741: gap of unknown length
* 14741 17016: contig of 2275 bp in length
* 17016 17116: gap of unknown length
* 17116 19133: contig of 2017 bp in length
* 19133 19233: gap of unknown length
* 19233 20261: contig of 1028 bp in length
* 20261 20361: gap of unknown length
* 20361 22138: contig of 1777 bp in length
* 22138 22238: gap of unknown length
* 22238 24389: contig of 2151 bp in length
* 24389 24489: gap of unknown length
* 24489 26393: contig of 1904 bp in length
* 26393 26493: gap of unknown length
* 26493 28677: contig of 2184 bp in length
* 28677 28777: gap of unknown length
* 28777 30246: contig of 1469 bp in length
* 30246 30346: gap of unknown length
* 30346 32008: contig of 1662 bp in length
* 32008 32108: gap of unknown length
* 32108 33927: contig of 1819 bp in length

33928 34027: gap of unknown length
* 34028 36660: contig of 2633 bp in length
* 36661 36760: gap of unknown length
* 36761 38570: contig of 1810 bp in length
* 38571 38671: gap of unknown length
* 38671 41138: contig of 2468 bp in length
* 41138 41238: gap of unknown length
* 41238 43167: contig of 1928 bp in length
* 43167 43267: gap of unknown length
* 43267 47194: contig of 3928 bp in length
* 47194 47294: gap of unknown length
* 47295 51037: contig of 3743 bp in length
* 51037 51137: gap of unknown length
* 51137 54101: contig of 2964 bp in length
* 54101 54201: gap of unknown length
* 54201 56590: contig of 2389 bp in length
* 56590 56691: gap of unknown length
* 56691 59953: contig of 3263 bp in length
* 59953 60054: gap of unknown length
* 60054 64339: contig of 4285 bp in length
* 64339 64438: gap of unknown length
* 64438 68405: contig of 3967 bp in length
* 68405 68505: gap of unknown length
* 68505 73010: contig of 4504 bp in length
* 73010 73109: gap of unknown length
* 73109 79117: contig of 6008 bp in length
* 79117 79217: gap of unknown length
* 79218 82852: contig of 3635 bp in length
* 82852 82953: gap of unknown length
* 82953 87619: contig of 4667 bp in length
* 87619 87720: gap of unknown length
* 87720 96467: contig of 8748 bp in length
* 96467 96567: gap of unknown length
* 96567 103197: contig of 6630 bp in length
* 103197 103297: gap of unknown length
* 103297 109986: contig of 6689 bp in length
* 109986 120973: gap of unknown length
* 120973 121073: contig of 10887 bp in length
* 121073 121074 130608: contig of 9535 bp in length
* 130608 130708: gap of unknown length
* 130708 143095: contig of 12387 bp in length
* 143095 143195: gap of unknown length
* 143195 153124: contig of 9929 bp in length
* 153124 153224: gap of unknown length
* 153224 170238: contig of 17014 bp in length.
* 153225 Location/Qualifiers

FEATURES
source 1. 170238
/organism="Rattus norvegicus"
/db_xref="taxon:10116"
/clone="CH230-297H21"

.BASE COUNT 44877 a 35872 c 35293 g 45438 t 8758 others
ORIGIN

Query Match 89.5% Score 17: DB 2: length 170238; 1
Best Local Similarity 100.0% Pred. No. 1.1le02;
Matches 17: Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 TGGGGCAGTATTGCA 18
Db 105814 TGGGGCAGTATTGCA 105798

RESULT 10
AK091176/c

LOCUS AK091176 2785 bp mRNA linear PRI 15-JUL-2002
DEFINITION Homo sapiens CDNA FLJ33857 f15, clone CTONG2006004.
ACCESSION AK091176
VERSION AK091176.1 GI:21749482

KEYWORDS oligo capping; f15 (full insert sequence).
SOURCE Homo sapiens tongue, tumor tissue CDNA to mRNA, clone_11b:CTONG2
ORGANISM Homo sapiens
Clone:CTONG2006004.

REFERENCE	AUTHORS	TITLE	JOURNAL	REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT	
1	Nishii, T., Oca, T., Nakagawa, S., Senoh, A., Mizuguchi, H., Inagaki, H., Sugiyama, T., Irie, R., Otsuki, T., Sato, H., Wakamatsu, A., Ishii, S., Yamamoto, J., Isono, Y., Kawai-Hio, Y., Saito, K., Nishikawa, T., Kikunaga, K., Yamashita, H., Matsuo, K., Nakamura, Y., Sekine, M., Kikuchi, H., Kanda, K., Magatsuma, M., Murekawa, K., Kanehori, K., Takahashi-Fujii, A., Oshima, A., Sugiyama, A., Kawakami, B., Sugano, S., Nagahari, K., Masuno, Y., Nagai, K. and Isogai, T.	Unpublished							
2	(Bases 1 to 2785)								
3	Isogai, T. and Yamamoto, J.								
4	Submitted (04-JUL-2002)								
5	Kazuo-Kamataari, Kitarazu, Chiba 292-0812, Japan								
6	(E-mail: genomics@hri.co.jp, Tel: 81-438-52-3975, Fax: 81-438-52-3986)								
7	NEDO human cDNA sequencing project supported by Ministry of Economy, Trade and Industry of Japan; cDNA full insert sequencing: Research Association for Biotechnology (RAB); cDNA library construction: Helix Research Institute (HRI) (supported by Japan Key Technology Center etc.); 5' and 3' end one pass sequencing: RAB, HRI, and Biotechnology Center, National Institute of Technology and Evaluation; clone selection for full insert sequencing: HRI and RAB; annotation: HRI and RAB.								
8	Location/Qualifiers								
9	1. 2785								
10	/organism="Homo sapiens"								
11	/db_xref="taxon:9606"								
12	/clone="CTONG2006004"								
13	/tissue_type="tongue, tumor tissue"								
14	/clone_id="CTONG2"								
15	/note="cloning vector: pME18SFL3"								
16	204..2630								
17	/note="unnamed protein product"								
18	/codon_start=1								
19	/protein_id="BAC03600.1"								
20	/db_xref="GI:21749483"								
21	/translation="MKVANIILSYKCNEDAOYFIREVAEKLKSLADGKETTTEEI LVLEEGLTMSIQOQELKRLSESSSOMALVYOCRLHNKMLSTSLRECKAKNDMLC FIHSGLAHNHPAEVKSLIOYFSPVIDHILAPENLPSVPTSKMDSPONOKRPOE OGSKQEWTDIPEILLOSSEPDSSWMLLVAVYKQAPILSYLACLOGAKASISCLCAV TITSVEDNATPEAKGHIODSTEDTHNLEDLSVILKTLTRKSKTILRGOLFCKDS PLLVMDNVEALCMFERNYKEAEKALLFEQKLETLNLAATVHPVTPAMVLEDVCFE LKLTMLQKQYVELGKLIQLEVEREHLEFSQDPVYKLCILCOIKLSDIAINHTITS YSLENTLOHCGSLILERLODTGOFALARVAELAPVDNLVYKETOBOMLKHEQV SLKAOADFPWKKHFNKKNKSSIKSAKSPSTQAVACHPOTGSMSEHEHLITLTLA GHMLAODVDPDKLELEKROIMCRTOHNGVACHPOTGSMSEHEHLITLTLA ASERSFSKLAALNTSKYLELNSLPSKETCEKRLDKMDESNLFLGLRLDCGCHEAR RVCTYFHFNPDAVLVLCRALASGEASMEDELHIALQASALEADIPRLRR HSTSLDSKFEVTVSSNEVNTNLEVTSKCIHKNKCRQVLCYDLAKELGCGSTVDG AAODGEMLKRIILASQOPDRCKRAQAFISTGLKPRDVAELVAEEVRELLTSSQGTG ALPFGDSQPL"								
22	BASE COUNT	800 a	624 c	643 g	718 t				
23	ORIGIN								
24	Query Match	86.3%	Score 16.4	DB 9	Length 2785				
25	Best Local Similarity	94.4%	Pred. No. 3e+02						
26	Matches 17	Conservative 0	Mismatches 1	Indels 0	Gaps 0				
27	DB	1	CTGGGGGCGAGTTATTGCA	18					
28	677	CTGGGGGCGACTTATTGCA	660						
29	RESULT 11								
30	AB058743	7752 bp	mRNA	linear	PRI 05-JUL-2002				
31	LOCUS	HOMO sapiens mRNA for KIAA1840 protein, partial cds.							
32	ACCESSION	AB058743							
33	VERSION	AB058743.2	GI:20521987						

[illegible]

ence)"
join(3516. .3573,3826.3921,7275.7495,7758.7874,
8133.8806,8857.9264,9413.9644,9693.10014,10075.10199,
10483.11115)
/gene="lin-35"
/note="C. elegans synthetic multivulva protein LIN-35
(GB:AF116529); coded for by the following C. elegans
cDNAs: yk13c4.3, yk13c4.5, yk251e2.3, yk251e2.5,
yk356a3.3, yk356a3.5, AF116529"
/codon_start=1
/product="C. elegans LIN-35 protein (corresponding
sequence C32F10.2)"
/protein_id="AAC24266.2"
/db_xref="GI:6671809"
/translation="MKRAADPEPGSTTDPFHEQSPDVAVLAGEFTTDTICEPPAKR
IDLIDKOEENGGVDSGLIKNESELTQMIKQETEBNIEAREEEDBEDQDSRTSM
PALLEDDEEDDADSGFLDKNTPTPPSQSPGCFANLPNDIVGAMETYNHAVOR
VLADESSEMSQSAIYYLLSLGSKRGRTIRLLIDPFPRICKITPEELIKVGSNED
KARVEEHSKSKIRRYOYIRIOGLGALVYKFFKRIKTEELIKVGSNEDPS
SHELETVLMTSLVAKSRRTVDLISNOLLESILDQVTEMCSMEGIVYHNHAKQFY
EDLENDCITKALCTOEGGSVLDARHESDHPFKMEKGTGIPSTNMFQERDLINVP
KTYAENLILQRSIDERITIPSEVDFSKTFQSPDTSVADILKVSYSGRFRAIELPT
KLSNHCKEKLALGVKASEKLVTSQSDRPPACVYNELGYSFQDLSSNLSLR
LKIKSNKMLEKLAEVCGTMSDSPAIIILKSDMTKKEFTTSLAEIGETINEMIP
KHYVARKELEIVPLIFEMKTIYAEIKVAREEDLIVIRBEEIDLSVFCFCEVILLV
NSGYARPPMSALGCVHPEMFHKVIDLITMTERKOLSRQWFKSEIEEYVIEFSK
SSSPLMPHVRNCPFAHFQFEGEDMAKLNSYSPKTFPIKKPDLEDELGRPIVQNO
TSRTKPIELFKRTYFAARLQDLTQSVSGARAKSQGWSLVFLLENDDLIFMDRIHL
QILLCVFIKRTKINIESMLFTEIMAYQRRQSANSLIYVGSVYFQDLSSNLSLR
KTIIELEKPGKEKTTVDIIKYNIIEFDRIKYIIGQIDSADDEDLMEWPATEESGL
MVRVYILFKLSLQIOLPKRKHGSKQERAIANLEKGIITIAMERSD"
complement(4284.5781)
/gene="C32F10.4"
/note="for a graphical representation of this gene see:
[www.wormbase.org/db/seq/sequence?name=C32F10.4;class=Sequ
ence]"
complement(join(4284.5781,4424.4733.4964,5009.5183,
5229.5322,5426.5605,5653.5781))
/gene="C32F10.4"
/note="coded for by the following C. elegans cDNAs:
yk150f2.3, yk150f2.5"
/codon_start=1
/product="Hypothetical protein C32F10.4"
/protein_id="AAC24267.1"
/db_xref="GI:1946999"
/translation="MTTESQTTDPYQHPQOARPEHPQPTORIAAEPEPFCGLPIGG
AGVIGSLIMICQIFGLIYSPFLVFEISIGSANDQDVATKACRKRTIATYAMVY
LSMTYFWPEWELIGAVIACIYLSDDNDNLTSCNOMQDVTAKCVKTELEMTI
SCFPLVSTIVALQKRYMILKNTFLRRNAYRLPQOASNTYNOYIVLPWMSGMP
PPRYTARAPMSQTAPEPIPTKQDPVDSSTTSASTSGSEFTPIITLVANSEPGA
DMSATINDPDRPRATSRPAYSVDVEEGHITVRSDVLO"
complement(11545.13911)
/gene="hmg-3"
/note="for a graphical representation of this gene see:
[www.wormbase.org/db/seq/sequence?name=C32F10.5;class=Sequ
ence]"
complement(join(11545.11796,11935.12877,12966.13603,
13675.13911))
/gene="hmg-3"
/note="Contains similarity to Pfam domain: PF00505
(HMG_box), Score=79.3, E-value=2.3e-20, N=1; coded for by
the following C. elegans cDNAs: cm1h10, yk20g10.5,
yk252d2.3, yk252d2.5, yk421h10.3, yk421h10.5"
/codon_start=1
/product="C. elegans HMG-3 protein (corresponding sequence
C32F10.5)"
/protein_id="AAC24268.1"
/db_xref="GI:1947000"
/translation="MTLELKFQGVYVEDIHLNGLTLTENSINFTIGDKGKSVYTG
TQVDRKQKQNLNRPGLRVLGSDGAGRGGFLDDDLQKOSTSSMSKSIQNSNLE
INGMNVGADVGNKRIEFSMENEPRIEICTQNSANTYANKNEEILFHEHQSRYOLM
EKREHNPDLNEDNEDTKYERKAVYLAAGLEAFEBOPICLLTDLICTTPRRRYDK
VPTSLAHGKGYDKIPVKTINRLVLVPHKDSROYVYLSLNPPIRQGGTHTSYLFE

Query Match	Best Local Similarity	86.3% 94.4%	Score 16.4; Pred. No. 2.7e+02;	DB 3;	Length 30102;
Matches	1/; Conservative	0;	Mismatches	1;	Indels 0; Gaps 0;
Oy	1	CTGGGGCAGATTATGCA	18		
Db	5289	CTGGGACAGATTATGCA	5272		
RESULT 13	AF003145/c	38525 bp	DNA	linear	INV 31-MAY-2002
LOCUS	Caenorhabditis elegans cosmid B0414, complete sequence.				
DEFINITION	AF003145				
ACCESSION	AF003145.1	GI:2088760			
VERSION	HTG.				
KEYWORDS	Caenorhabditis elegans.				
SOURCE	Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabdilitida;				
ORGANISM	Rhabdilitida; Rhabdilitidae; Pelodierinae; Caenorhabditis.				
REFERENCE	1 (bases 1 to 38525)				
AUTHORS	Waterston, R.				
TITLE	Genome sequence of the nematode C. elegans: a platform for				
JOURNAL	investigating biology. The C. elegans Sequencing Consortium				
MEDLINE	Science 282 (5396), 2012-2018 (1998)				
PUBMED	99069613				
REFERENCE	2 (bases 1 to 38525)				
AUTHORS	Sammons, L., Mohlmann, P. and Rohlfing, T.				
TITLE	The sequence of C. elegans cosmid B0414				
JOURNAL	Unpublished (2001)				
REFERENCE	3 (bases 1 to 38525)				
AUTHORS	Waterston, R.				
TITLE	Direct Submission				
JOURNAL	Submitted (08-MAY-1997) Department of Genetics, Washington				
REFERENCE	University, 4444 Forest Park Avenue, St. Louis, MO 63108, USA				
AUTHORS	4 (bases 1 to 38525)				
TITLE	Waterston, R.				
JOURNAL	Direct Submission				
REFERENCE	Submitted (05-JAN-2000) Department of Genetics, Washington				
AUTHORS	University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA				
TITLE	5 (bases 1 to 38525)				
JOURNAL	Waterston, R.				
REFERENCE	Direct Submission				
AUTHORS	Submitted (28-JUN-2001) Department of Genetics, Washington				
TITLE	University, Genome Sequencing Center, 4444 Forest Park Avenue, St.				
JOURNAL	Louis, MO 63110, USA				
REFERENCE	6 (bases 1 to 38525)				
AUTHORS	Waterston, R.				
TITLE	Direct Submission				
JOURNAL	Submitted (19-APR-2002) Department of Genetics, Washington				
REFERENCE	University, Genome Sequencing Center, 4444 Forest Park Avenue, St.				
AUTHORS	Louis, MO 63110, USA				
TITLE	7 (bases 1 to 38525)				
JOURNAL	Waterston, R.				
REFERENCE	Direct Submission				
AUTHORS	Submitted (31-MAY-2002) Department of Genetics, Washington				
TITLE	University, Genome Sequencing Center, 4444 Forest Park Avenue, St.				
JOURNAL	Louis, MO 63110, USA				
COMMENT	Submitted by: Genome Sequencing Center Department of Genetics, Washington University St. Louis, MO 63110, USA, and Sanger Centre, Hinxton Hall Cambridge CB10 1RQ, England email: tw@nematoide.wustl.edu and jes@sanger.ac.uk				

NOTICE: This sequence may not be the entire insert of this clone. It may be shorter because we only sequence overlapping sections once, or longer because we provide a small overlap between neighboring submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one m13 subclone.

For a graphical representation of this cosmid sequence and its analysis see:
<http://www.wormbase.org/db/seq/sequence?name=B0414.3;class=Sequence>

NEIGHBORING COSMID INFORMATION

The 5' cosmid is T22E7, 2000 bp overlap; the 3' cosmid is C32F10, 5900 bp overlap. Actual start of this cosmid is at base position 1 of B0414; actual end is at 7060 of C32F10.

NOTES:

Coding sequences below are the result of integration and manual review of the following data: computer analysis using the program GeneFinder (P. Green and L. Hillier, personal communication), the large scale EST projects of Yujii Kohara (http://www.ddbj.nig.ac.jp/c-elegans/html/CE_INDEX.html) and The C. elegans ORFome cloning project (<http://worldb.dfci.harvard.edu/>), similarity to other proteins from Blastx analyses (<http://blast.wustl.edu/>), sequence conservation with C. briggsae using Jim Kent's WABA alignment program (Genome Research 10:1115-1125, 2000), individual C. elegans Genbank submissions, and personal communications with C. elegans researchers. tRNAs are predicted using the program tRNAscan-SE (Lowe, T.M. and Eddy, S.R., 1997, Nucl. Acids. Res., 25, 955-964).

Location/Qualifiers

1. .38525

/organism="Caenorhabditis elegans"

/strain="Bristol N2"

/db_xref="taxon:6239"

/chromosome="I"

/clone="B0414"

.925. .1308

/gene="B0414.4"

/note="for a graphical representation of this gene see: <http://www.wormbase.org/db/seq/sequence?name=B0414.4;class=Sequence>"

CDS

join(925..1086,1147..1308)

/gene="B0414.4"

/codon_start=1

/product="Hypothetical protein B0414.4"

/protein_id="AAB57716.1"

/db_xref="GI:2088763"

/translation="MNGLSVNTICQIPADIDPFLPSSLSLISNNPITCCATRLTSSDDVFLNQADTKCASPPELGKRIFFELSDYCTTRNPKKASFFQFLVFLVAVGSEFLP"

3032. .3868

/gene="B0414.3"

/note="for a graphical representation of this gene see: <http://www.wormbase.org/db/seq/sequence?name=B0414.3;class=Sequence>"

gene

join(3032..3268,3428..3868)

/gene="B0414.3"

/note="Contains similarity to Pfam domain: PF00538 (linker histone), Score=115.3, E-value=3.6e-31, N=1, coded for by the following C. elegans cDNAs: yk67e4.5, yk77a3.5, yk77a3.3, yk79e10.5, yk95h6.3, yk95h6.5, yk106g6.3, yk106g6.5, yk114a2.3, yk114a2.5, yk125e1.5, yk125e1.3, yk151a6.5, yk151a6.3, yk172d2.3, yk172d2.5, yk173g2.3, yk173g2.5, yk184c6.3, yk184c6.5, CEESX29F, CEMS55F,

CDS

gene

complement(12252..16800)
/product="Hypothetical protein B0414.3"
/protein_id="AAB57720.1"
/db_xref="GI:2088767"
/translation="MSDVVAEPPAVKPTKASKATKAKATKIPKVVAAAPPFLMTEAVSNLDRKSSRVAFKFTLAKYTGDDVNTNALRSALKKGVYSLVOTNGICANGFRVAVAKKPAVKATGCKWKVTKAKKASVSDKAKTKATGCKVKKVSPKRIKPAVKVTKAAAPTRSAANETAPKAAETAAAPKAAVTKATKTPARKA VGTAPKA"
12252. .16800
/gene="rnt-1"
/note="for a graphical representation of this gene see: <http://www.wormbase.org/db/seq/sequence?name=B0414.2;class=Sequence>"
join(12252..13320,13218..13335,13592..13711,14513..14586,15563..15679,16289..16347,16400..16518,16571..16800)
/gene="rnt-1"
/note="C. elegans homolog for Drosophila rnt (RNT-1); contains similarity to Pfam domain PF00853, Score=159.4, E-value=1.6e-44, N=1, coded for by the following C. elegans cDNAs: yk305f9.3, yk305f9.5, AB027412"
/codon_start=1
/product="C. elegans RNT-1 protein (corresponding sequence B0414.2)"
/protein_id="AAB57715.2"
/db_xref="GI:6671807"
/translation="MTNVEHVRNFTIEQOPAPAKTLKSSNNILYALPKMRSKSFEPEPVVLLTPVPDNTVEVSIWAGDERKECEVREKAVHVOAKFNDLRVSGRGRKFLHTIYHAPMVAWVAVKIVAVGVPDARIPKQGSILKQAEOQTLEPDDIRPQGPMPMTMTIPMPFPLPMTQTEPPSPFISGPHSISALAMKIHSEMKPIKOKVEDNRVLTSSNCSISFRTPSDQRKLKRRSPRSITKSETSLNLTQETPESVESKRRNVSTSSSSSPTIWRP"
complement(17581..20134)
/gene="cpb-3"
/note="for a graphical representation of this gene see: <http://www.wormbase.org/db/seq/sequence?name=B0414.5;class=Sequence>"
complement(join(17581..17792,17868..18080,18157..18757,18814..18980,19035..19964,20012..20134))
/gene="cpb-3"
/note="strong similarity to Xenopus cytoplasmic polyadenylation element binding protein (MID:9624634); coded for by the following C. elegans cDNAs: yk88f12.3, yk88f12.5, yk220a12.3, yk220a12.5, yk240d6.5"
/codon_start=1
/product="C. elegans CPB-3 protein (corresponding sequence B0414.5)"
/protein_id="AAB57717.1"
/db_xref="GI:2088764"
/translation="MNLNDRVEIGENRISIVAKSTERRKNIIPSLKPLSPVESVGEKSPTPAVYDLFEKKYKSDSVKPTREDMMNNGFDLSIDENKGFRLKLOMLVGSKSKSYDESPSPANRLKLTVPSPRTSTEHPSNRRHIGSSSTTKTSDSENSQRKKSARLNFEDRQKDKSTRTKNSSTFAVETRETYIVTSANESGAPVIRKANRKLINPKGSDLEPTSPDQGFSSSTESSPSDSMNQFPPSEHRTSANESGAPVIRKANRKLINPKGSDATMFGELPRPDYTPSPSRKIFVGGVPMIDTAAKDSGEGSCAVEPQGEARYRS GSNLAPNSNGSLAQTGTGAATGYVVMIDFEDRAVVASILHECQGLGAEVAFKIRBORSKSTEIRVOGOPOYVOIIPWTSDSMSEDDTLTLETGIEPRKTVFVGAHLGHTM AOVLISINEDCGAYECVQOLDTRKFKYIPGSRVYRFRGATGAFKALIEGDIHYHSKFKRVOIDPLESTCCWVCNSQSAHCFRCRNKFKYICHNCALDGGKDDCVHILPVI VPSASAKAFGATRRSQVSSLSLKNNSNNSQVNNLSIPHPSAGFPIIVGAPQTL S ALVGYLQNGOQLKPAVEFPWTPSPSEINKNRRSFTEPPTVTFPSTVISPQKS VSDGSPFLPAYVNSAFLPPTPYGSPNHPSTSNISQPOQPYGANLYGFLPQOHP SAMRSPNVGQ"
complement(20992..23452)
/gene="gln-3"
/note="for a graphical representation of this gene see: <http://www.wormbase.org/db/seq/sequence?name=B0414.6;class=Sequence>"

CDS

gene

CDS

gene

gene

Query Match	86.38;	Score 16.4;	DB 3;	Length 38525;
Best Local Similarity	94.48;	Pred. NO. 2.7e+02;		
Matches 17; Conservative	0;	Mismatches 1;	Indels 0;	Gaps 0;

QY 1 CTGGGGCAGTTATGCA 18

Db 37914 CTGGAGCAGTTATTGCA 37897

RESULT 14

LOCUS	42325 bp	DNA	linear	BCT 01-MAR-1994
DEFINITION	Mycobacterium leprae cosmid B1620.			
ACCESSION	U00015			
VERSION	U00015			

VERSION	U00015.1	GI:466931
KEYWORDS		
SOURCE	Mycobacterium leprae.	
ORGANISM	Mycobacterium leprae	

REFERENCE
AUTHORS
JOURNAL
REFERENCE
AUTHORS

1 (bases 1 to 42325)
Smith, D. R.
Unpublished
2 (bases 38281 to 41190)
Rinke de Wit, T. F., Nabeije, S., Oslund, A., Mito, T. I., Hermans, D. W.

TITLE	JOURNAL	YEAR	PAGE
Sequence and immunological characterization of a serine-rich antigen from <i>Mycobacterium leprae</i>	Infect. Immun.	61 (5)	2145-2153 (1993)

FEATURES

Location/Qualifiers

```

/transl_table=11
/product="Otp"
/protein_id="AAC43238.1"
/db_xref="GI:466950"

```

```

/note="weak match 42.1 KD protein; B1620_C3_270"
/codon_start=1
/transl_table=1
/product="u1620m"
/protein_18="AACCA3237.1"
/db_xref="GI:466949"
/translation="MVCILRLAGLEKLPILVLSRYGVGALRVLELGLAPINISITY
LTGVSRPADLDLEENRADGVSY"
complement(2282..2473)

```

```

/note="45 KDa serine-rich antigen; B160_C1_177"
/codon_start=1
//transl_table=1
//product="ag45"
//protein_id="AAC43220.1"
/db_xref="GI:466932"
//translation="MFDMVYSPEYNAFLMRGSGPTPLMGAAEMATISLABOLEAAQ
AGPDLIVAVASPAFGESDMLASRVSFFVAILDGMNENGLITRVYLAHAYVAFEEFL
AGMPDLIVLGNLIHTMALAKLNNFGOVSTYVAALADYDNLQNSTAMTYTIDTYL
RETKMNEFEPAPOLVSRRCMDRDRDSVNSFFSSSSDSLEYSIDNLVDSVAQSEHDS
DTMSQSVSCVSAQSELCDSFPGPSQSSNDLSAKISLQQLGGDLSLISASASIS
LTTNLSISSTTSSIMPIYASQVETFLGSGVAYEACISSTFASVDVAASKVAVCV
GOAIVSGGLARVPENMATSQPVMAFHSVPAGCAITTAVSGPLEGYQPREVLTAAS
VAGSGTGCPAFNEAY"
complement(5262..5564)

```

```

/note="weak match 32KD protein (sp:p16645); B1620_C2_218
/codon_start=1
/transl_table=11
/product="p1620k"
/protein_id="AAC4325.1"
/db_xref="GI:166937"
/translation="MSFPLRVEVGLIMAAGRLERTISESACNAKLTPLYTKVPPPA
ADQVKLVISQVFSSSGKGYEGYAAQGVDSRLFVQSLKDAAGYDMSDHYMLNTD"
6340..6816

```

```

/note="transposase; B1620_F1_4"
/codon_start=1
/translation_table=1
/product="u1620a"
/protein_id="AAC43241.1"
/db_xref="GI:465953"
/translation="MEDICVANRRVGYVKRLRNTRAKRGDHRDHIDIGVARNWICLAG
IDGIVYRDRITRTTEHDKQAPRGHRYVDPQELVNAKPGSVESPIISMFRRCRGCT
TAFCDVSRNIPGPGRTSTVLVTSLEQALFIRKTYELRFAMDESAPLCCR
complement(7286..7606)

```

```

/code="transposase: B1620_03_266"
/codon_start=1
/transl_table=1
/product="u1620b"
/protein_id="AAC4235.1"
/db_xref="GI:466947"
/translation="MVNVALITTDVNAEGYRETLGGCCDHRRRRRGLVEVSAADRR
LAFRRYGHORCRPRITLRHNOGQPPGVFMORCIRYILGHDPHGIRLTHGFLVVGAPYPT

```

CDS
ILGV"
complement(8447. .9097)
/codon_start=1
/transl_table=11
/product="B1620_C3_264"
/protein_id="AAC43234.1"
/db_xref="GI:466946"
/translation="MGRIRIRRLPMPSTPAPQEFRRDVVVVEPGNDPDRDCDTP
DMGVGAAMHTIDPILVGFSSASQDQRTDYNRGQAEKQKEPVALFEYTNP
PRTVSVYLMGDRIDHVELSKRTVMSSQGVSEILYADLAKRQASQVETIDKL
SOLADDERHVALRESVGNNTNLPSEQAALAEAEVATRTSDICPADTNDOM"
complement(9111. .9431)
/codon_start=1
/transl_table=11
/product="B1620_C2_214"
/protein_id="AAC43224.1"
/db_xref="GI:466936"
/translation="MRSKIDNTVOSEHINSIASQEHENPAACASSGVSAAGLANVS
TSHGSCAOFNDTLKMYEDAHHTLDESLHTGGIDLARVLRAAAMVCDADETCGSDIK
SAFC"
complement(9489. .10673)
/codon_start=1
/transl_table=11
/product="B1620_C2_213"
/protein_id="AAC43223.1"
/db_xref="GI:466935"
/translation="MSGAFIIDPTLKATEAMHALLIGVNDGVYLSLSLFFERALE
HIAAFPGDGMGSADKAGQNRKRVDFQELAEIDKELILHNOANSVOTGIL
DKAKALLEPVAIDILNIPVGSVMSIOACAAAMAASGGLAYLIVOTPIHT
AKFVALLIAHLASAVADYVSDGVAITIKGIVDHMHFAGALIGIKIVKTIHMF
PELFHMSRHSFPGGIPGLSGATSGLSQVYTGICVPLAASSGGLISESLSTENT
PSLAVGAGLGLSLPQLAQHLAASRQSTQAGVSAELSTEGFGQDEPVSAGCSO
GMSGGQMGMTASTSKSKDKRRKKRYSEGAAGTDAERAPIEVQSGGRALAOH
VY"
CDS
13956
/codon_start=1
/transl_table=11
/product="B1620_F1_14"
/protein_id="AAC43239.1"
/db_xref="GI:466951"
/translation="MSKPLQPCADHIAAVLFGMTNFOAMNFGESPVAIGPAHQAK
AKVGVKPRCALTNISILPTVTQTTSMWPRREQLRPIALS"
16793. .17880
/note="D-Ribose periplasmic binding protein; B1620_F2_67"
/codon_start=1
/transl_table=11
/product="rdsb"
/protein_id="AAC43242.1"
/db_xref="GI:466954"
/translation="MSTGTSVPLRVASIMLILGLLPALIPACGFSESHHKLTISYP
TANSPFWMAVYRIDEQSHQGLVYINAVSAEEDQKQSDVETLLNOGDLIVTPQS
TSIAPTLRIATQANIPIVYVDRPGYAFQGNKNADYVAFGLPNNKESGSAEALMA
GGSKFALIGAPGNSVAGRKAGLESALTAGHRLVQFQVQDSEKSLAEENLMO
AHPADGNAIWCENDKICOGATIKAVYNNRKEKPIEGGMDLIPQALAIENELIYVSL
GAHMLEGGGGLAILYRKIHGDPALERMVYKLDLKYDKSVNAFKARYIDNTRHYMKS
TTPEDMTITLN"
18218. .19951
/codon_start=1
/transl_table=11
/product="B1620_F2_68"
/protein_id="AAC43243.1"
/db_xref="GI:466955"
/translation="MVIDIGRIESINPATRILGLRANVDYMKHGHPFCFYDTONR
VDLEREVVYRREVTYSKYVIGDQHSQRLMSVNSILAYKAPPHSALVYFSDI
SAHHSIERLTVEATVHDCITGLANRRPAEDQITKSLQDERSRLAAVILLDDPKVI

Query Match 86.3%; Score 16.4; DB 1; Length 42325;
Best Local Similarity 94.4%; Pred. No. 2.7e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 CTGGGGGCGAGTTATGCA 18
|||||
DB 29336 CTGGAGGCGAGTTATGCA 29319

RESULT 15
AC101474
LOCUS
DEFINITION
AC101474
ACCESSION
AC101474
VERSION
AC101474.1
KEYWORDS
HTG: HTGS, PHASE0.
SOURCE
Mus musculus
ORGANISM
Mus musculus
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
AC101474 65707 bp DNA linear HTG 23-NOV-2001
Mus musculus clone RP23-187L16, LOW-PASS SEQUENCE SAMPLING.
AC101474
AC101474.1 GI:17060249
HTG: HTGS, PHASE0.
Mus musculus.
Mus musculus.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 65707)
Birren, B., Linton, L., Nusbaum, C. and Lander, E.
Mus musculus, clone RP23-187L16
Unpublished
2 (bases 1 to 65707)
Birren, B., Linton, L., Nusbaum, C., Lander, E., All, A., Allen, N.,
Anderson, S., Barna, N., Bastien, V., Boguslavsky, L., Boukhgalter, B.,
Brown, A., Camarata, J., Campopiano, A., Chang, J., Chazaro, B.,
Chospel, Y., Colangelo, M., Collins, S., Collamore, A., Cook, A.,
Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S., Dodge, S., Faro, S.,
Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Garbaya, S.,
Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N.,
Hages, B., Heaford, A., Horton, L., Hulme, W., Illey, I., Johnson, R.,
Jones, C., Kamat, A., Karatas, A., Kells, C., LaRoque, K.,
Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Liu, G.,
Maclean, C., MacDonald, P., Major, J., Margulis, N., Matthews, C.,
McCarthy, M., McEwan, P., McKernan, K., McPeckers, R., Meldrum, J.,
Merru, C., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C.,
Norus, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D.,
Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V.,
Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P.,
Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupbach, R.,
Seaman, S., Severy, P., Spencer, B., Stenge-Thomann, N., Stojanovic, N.,
Struss, N., Subramanian, A., Talamas, J., Testaye, S., Theodore, J.,
Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H.,
Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G.,
Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.
Direct Submission
Submitted (23-NOV-2001) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
Project Information
Center project name: L16819
Center clone name: 187_L_16
NOTE: This record contains 82 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
1
725 824: gap of 100 bp
825 1548: contig of 724 bp in length
1549 1648: gap of 100 bp
1649 2372: contig of 724 bp in length
2373 2472: gap of 100 bp
2473 3152: contig of 680 bp in length

```
* 3153 3252: gap of 100 bp
* 3253 3923: contig of 671 bp in length
* 3924 4023: gap of 100 bp
* 4024 4694: contig of 671 bp in length
* 4695 4794: gap of 100 bp
* 4795 5497: contig of 703 bp in length
* 5498 5597: gap of 100 bp
* 5598 6307: contig of 710 bp in length
* 6308 6407: gap of 100 bp
* 6408 7120: contig of 713 bp in length
* 7121 7220: gap of 100 bp
* 7221 7935: contig of 715 bp in length
* 7936 8035: gap of 100 bp
* 8036 8760: contig of 725 bp in length
* 8761 8860: gap of 100 bp
* 8861 9589: contig of 729 bp in length
* 9590 9689: gap of 100 bp
* 9690 10407: contig of 718 bp in length
* 10408 10507: gap of 100 bp
* 10508 11190: contig of 683 bp in length
* 11191 11290: gap of 100 bp
* 11291 11966: contig of 676 bp in length
* 11967 12066: gap of 100 bp
* 12067 12763: contig of 697 bp in length
* 12764 12863: gap of 100 bp
* 12864 13576: contig of 713 bp in length
* 13577 13676: gap of 100 bp
* 13677 14385: contig of 709 bp in length
* 14386 14485: gap of 100 bp
* 14486 15193: contig of 708 bp in length
* 15194 15293: gap of 100 bp
* 15294 16002: contig of 709 bp in length
* 16003 16102: gap of 100 bp
* 16103 16820: contig of 718 bp in length
* 16821 16920: gap of 100 bp
* 16921 17606: contig of 686 bp in length
* 17607 17706: gap of 100 bp
* 17707 18431: contig of 725 bp in length
* 18432 18531: gap of 100 bp
* 18532 19258: contig of 727 bp in length
* 19259 19358: gap of 100 bp
* 19359 20048: contig of 690 bp in length
* 20049 20148: gap of 100 bp
* 20149 20838: contig of 690 bp in length
* 20839 20938: gap of 100 bp
* 20939 21605: contig of 667 bp in length
* 21606 21705: gap of 100 bp
* 21706 22405: contig of 700 bp in length
* 22406 22505: gap of 100 bp
* 22506 23202: contig of 697 bp in length
* 23203 23302: gap of 100 bp
* 23303 24003: contig of 701 bp in length
* 24004 24103: gap of 100 bp
* 24104 24816: contig of 713 bp in length
* 24817 24916: gap of 100 bp
* 24917 25611: contig of 695 bp in length
* 25612 25711: gap of 100 bp
* 25712 26440: contig of 729 bp in length
* 26441 26540: gap of 100 bp
* 26541 27232: contig of 692 bp in length
* 27233 27332: gap of 100 bp
* 27333 27980: contig of 648 bp in length
* 27981 28080: gap of 100 bp
* 28081 28746: contig of 666 bp in length
* 28747 28846: gap of 100 bp
* 28847 29552: contig of 706 bp in length
* 29553 29652: gap of 100 bp
* 29653 30358: contig of 706 bp in length
* 30359 30458: gap of 100 bp
* 30459 31159: contig of 701 bp in length
* 31160 31259: gap of 100 bp
* 31260 31974: contig of 715 bp in length
* 31975 32074: gap of 100 bp
```

```
* 32075 32804: contig of 730 bp in length
* 32805 32904: gap of 100 bp
* 32905 33578: contig of 674 bp in length
* 33579 33678: gap of 100 bp
* 33679 34377: contig of 699 bp in length
* 34378 34477: gap of 100 bp
* 34478 35204: contig of 727 bp in length
* 35205 35304: gap of 100 bp
* 35305 35986: contig of 682 bp in length
* 35987 36086: gap of 100 bp
* 36087 36759: contig of 673 bp in length
* 36760 36859: gap of 100 bp
* 36860 37527: contig of 668 bp in length
* 37528 37627: gap of 100 bp
* 37628 38331: contig of 704 bp in length
* 38332 38431: gap of 100 bp
* 38432 39144: contig of 713 bp in length
* 39145 39244: gap of 100 bp
* 39245 39950: contig of 706 bp in length
* 39951 40050: gap of 100 bp
* 40051 40776: contig of 726 bp in length
* 40777 40876: gap of 100 bp
* 40877 41590: contig of 714 bp in length
* 41591 41690: gap of 100 bp
* 41691 42408: contig of 718 bp in length
* 42409 42508: gap of 100 bp
* 42509 43186: contig of 678 bp in length
* 43187 43286: gap of 100 bp
* 43287 43949: contig of 663 bp in length
* 43950 44049: gap of 100 bp
* 44050 44750: contig of 701 bp in length
* 44751 44850: gap of 100 bp
* 44851 45550: contig of 700 bp in length
* 45551 45650: gap of 100 bp
* 45651 46325: contig of 675 bp in length
* 46326 46425: gap of 100 bp
* 46426 47139: contig of 714 bp in length
* 47140 47239: gap of 100 bp
* 47240 47963: contig of 724 bp in length
* 47964 48063: gap of 100 bp
* 48064 48786: contig of 723 bp in length
* 48787 48886: gap of 100 bp
* 48887 49612: contig of 726 bp in length
* 49613 49712: gap of 100 bp
* 49713 50434: contig of 722 bp in length
* 50435 50534: gap of 100 bp
* 50535 51223: contig of 689 bp in length
* 51224 51323: gap of 100 bp
* 51324 51998: contig of 675 bp in length
* 51999 52098: gap of 100 bp
* 52099 52776: contig of 678 bp in length
* 52777 52876: gap of 100 bp
* 52877 53580: contig of 704 bp in length
* 53581 53680: gap of 100 bp
* 53681 54382: contig of 702 bp in length
* 54383 54482: gap of 100 bp
* 54483 55186: contig of 704 bp in length
* 55187 55286: gap of 100 bp
```

Query Match 86.3%; Score 16.4; DB 2; Length 65707;
Best Local Similarity 94.4%; Pred. No. 2.6e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 CTGGGGCAGTTATTCGA 18
DB 34780 CTGGGAGCAGTTATTCGA 34797

Search completed: July 1, 2003, 07:23:04
Job time : 363.359 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 1, 2003, 03:06:29 ; Search time 105.231 Seconds
(without alignments)
406.611 Million cell updates/sec

Title: US-10-053-662A-30

Perfect score: 19

Sequence: 1 ctgagggcagctatgacac 19

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 2185239 segs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_101002:*

1: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:*
2: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:*
3: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:*
4: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:*
5: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:*
6: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:*
7: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:*
8: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:*
9: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:*
10: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:*
11: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:*
12: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:*
13: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:*
14: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:*
15: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:*
16: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:*
17: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:*
18: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:*
19: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:*
20: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:*
21: /SID2/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:*
22: /SID2/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:*
23: /SID2/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:*
24: /SID2/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result NO.	Score	Query Match	Length	ID	Description
1	16.4	86.3	271	14	AA060904
2	15.8	83.2	1095	22	AAS29516
3	15.8	83.2	1183	22	ABA06384
4	15.8	83.2	1183	22	AAS29640
5	15.8	83.2	1183	22	AA163963
6	15.8	83.2	1183	22	AAS31277
7	15.8	83.2	1183	22	AAS34773
8	15.8	83.2	1183	24	ABQ6601
9	15.8	83.2	1490	20	AA52932

C 10	15.8	83.2	2613	22	AA163904	Human polynucleoti
C 11	15.8	83.2	3463	21	AA877382	Human OREF ORF2937
C 12	15.8	83.2	3531	23	AA568056	DNA encoding novel
C 13	15.8	83.2	6230	22	AA542520	Human cDNA encodin
C 14	15.4	81.1	164	23	ABV12689	Human prostate exp
C 15	15.4	81.1	427	23	ABV33819	Human prostate exp
C 16	15.4	81.1	427	23	ABV42711	Human prostate exp
C 17	15.4	81.1	3582	24	AAD31145	Human laminin gamm
C 18	15.4	81.1	3620	21	AA037332	Human laminin 5 cD
C 19	15.4	81.1	3720	21	AA037332	Human laminin 5 cD
C 20	15.4	81.1	4316	17	AA113324	Kallitn/laminin 5
C 21	15.4	81.1	4316	24	AA142911	Laminin gamma-2 ch
C 22	15.4	81.1	5020	21	AA037330	Human laminin 5 cD
C 23	15.4	81.1	5152	20	AA224627	Human lung tumor a
C 24	15.4	81.1	5156	21	AA055866	Human lung cancer-
C 25	15.4	81.1	5156	24	AB149085	Human lung tumour
C 26	15.4	81.1	5200	17	AA113323	Kallitn/laminin 5
C 27	15.4	81.1	5200	21	AA037329	Human laminin 5 cD
C 28	15.4	81.1	5200	21	AA142910	Laminin gamma-2 ch
C 29	15.4	81.1	9542	23	ABK42922	Genomic sequence #
C 30	15.4	81.1	11443	19	AAV52182	Streptococcus pneu
C 31	14.8	77.9	403	22	AA185494	Human polynucleoti
C 32	14.8	77.9	471	22	AA114658	Probe #4591 for ge
C 33	14.8	77.9	471	22	AA104468	Probe #4459 used t
C 34	14.8	77.9	475	24	AB140436	Sunflower lipid tr
C 35	14.8	77.9	566	22	ABA63470	Human foetal liver
C 36	14.8	77.9	566	22	ABA30668	Probe #9134 for ge
C 37	14.8	77.9	566	22	AAK12003	Human brain expres
C 38	14.8	77.9	566	22	AAK37709	Human bone marrow
C 39	14.8	77.9	566	22	AA118465	Probe #8398 for ge
C 40	14.8	77.9	566	22	AA13582	Probe #12268 used
C 41	14.8	77.9	566	24	AB111699	Human genome-deriv
C 42	14.8	77.9	694	22	AA137202	Human musculoskele
C 43	14.8	77.9	694	22	AA137204	Human musculoskele
C 44	14.8	77.9	867	21	AA053587	Streptococcus pneu
C 45	14.8	77.9	1213	21	AA13700	Aspergillus oryzae

ALIGNMENTS

RESULT 1	AA060904	AA060904 standard; DNA; 271 BP.
ID	AA060904;	
AC	AA060904;	
XX		
DT	16-MAR-1994 (first entry)	
XX		
DE	Human brain Expressed Sequence Tag EST000995.	
XX		
KW	Gene transcription product; genetic markers; tagging; in vivo;	
KW	transcription; mapping; locations; chromosomes; chromosomal; ss.	
XX		
OS	Homo sapiens.	
XX		
PN	W09316178-A.	
XX		
PD	19-AUG-1993.	
XX		
PF	12-FEB-1993;	93WO-US01294.
XX		
PR	12-FEB-1992;	92US-0837195.
XX		
PA	(USSH) US DEPT HEALTH & HUMAN SERVICE.	
XX		
PI	Adams MD, Moreno RF, Venter CJ;	
XX		
DR	WPI, 1993-272882/34.	
XX		
PT	Enriched oligonucleotides and corresp. sequences - used as	
PT	markers for human genes transcribed in-vivo, facilitate tagging	
PT	of most human genes	

XX PS Example 4; Page 409; 500bp; English.
XX CC The Expressed Sequence Tag was isolated from a human brain cDNA
CC library as part of a large set of ESTs which can be used as markers
CC for human genes transcribed in vivo. They can be used to facilitate
CC tagging of most human genes, for mapping locations of expressed genes
CC on chromosomes, for individual or forensic identification, for mapping
CC locations of disease-associated genes, for identification of tissue
CC type, and for prep. of antisense sequences, probes and constructs.
CC EST00995 has a "poor" coding probability as evaluated using the
CC coding-region prediction program CRM. See also AA039041-061440.
XX SQ Sequence 271 BP; 54 A; 98 C; 57 G; 60 T; 2 other;
QY Query Match 86.3%; Score 16.4; DB 14; Length 271;
Best Local Similarity 89.5%; Pred. No. 55;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Db 1 CTGGGGGCGAGTTATTGCAC 19
218 CTGGGGGCGTGTATTTCAC 236
RESULT 2
AAS29516/c
ID AAS29516 standard; cDNA; 1095 BP.
XX AC AAS29516;
XX DT 21-NOV-2001 (first entry)
XX DE Human endocrine polypeptide encoding cDNA SEQ ID No 16.
XX KW Endocrine protein; human; mouse; rabbit; goat; horse; food additive;
KW cat; dog; chicken; sheep; immunosuppressive; antiarthritic; vasotropic;
KW antirheumatic; antiproliferative; cytostatic; cardiac; neuroprotective;
KW cerebroprotective; noctropi; antibacterial; vitucide; fungicide; cancer;
KW ophthalmological; vulnetaire; gene therapy; autoimmune disease; neoplasm;
KW hyperproliferative disorder; breast; liver; cardiovascular disorder; ss;
KW cerebrovascular disorder; nervous system disorder; bacterial infection;
KW fungal infection; viral infection; ocular disorder; endocrine disorder;
KW gastrointestinal disorder; renal disorder; respiratory disorder;
KW wound healing; skin aging; organ transplantation; food preservative;
KW tissue regeneration; anti-fertility.
XX OS Homo sapiens.
XX PN MO200155364-A2.
XX PD 02-AUG-2001.
XX PF 17-JAN-2001; 2001MO-US01308.
XX PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.

PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225477.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 14-AUG-2000; 2000US-0226279.
PR 18-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226682.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234597.
PR 25-SEP-2000; 2000US-0234598.
PR 25-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.

PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246509.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Barash SC, Ruben SM;
XX
XX WPI: 2001-451936/48.
XX P-PSDB: AAU18287.
XX
XX Isolated polypeptide for treating, preventing and/or prognosing
XX disorders of the endocrine system such as reproductive disorders,
XX endocrine cancers and also for testing and detection e.g. diagnosis -
PS
PS Claim 1: SEQ ID No 16; 604pp; English.
XX
XX Sequences AAS29511-AAS29736 represent cDNA molecules, which encode the
XX endocrine polypeptides of the invention. Endocrine polypeptides and their
XX associated polynucleotides of the invention are useful in the diagnosis,
XX treatment and prevention of various types of disorders in e.g. humans,
XX mice, rabbits, goats, horses, cats, dogs, chickens or sheep. A
XX pathological condition can be determined by determining the presence or
XX absence of a mutation in an endocrine polynucleotide. The treatable
XX disorders include autoimmune diseases such as rheumatoid arthritis,
XX hyperproliferative disorders such as neoplasms of the breast or liver,
XX cardiovascular disorders such as cardiac arrest, cerebrovascular
XX disorders such as cerebral ischaemia, nervous system disorders such as
XX Alzheimer's disease, infections caused by bacteria, viruses and fungi,
XX ocular disorders such as corneal infection, endocrine disorders such as
XX premature labour and infertility, gastrointestinal disorders such as
XX Crohn's disease, renal disorders such as glomerulonephritis and

CC respiratory disorders such as asthma. The polypeptides can also be used
CC to aid wound healing, to prevent skin aging due to sunburn, to maintain
CC organs before transplantation, to regenerate tissues and in chemotaxis.
CC The polypeptides can also be used as a food additive or preservative to
CC increase or decrease storage capabilities.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO

Query Match 83.2%; Score 15.8; DB 22; Length 1095;
Best Local Similarity 89.5%; Pred. No. 1.3e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 CTGGGCGAGTTATTCAC 19
||||||| ||||| |||
Db 663 CTGGGCGCTGTATTCAC 645

RESULT 3
ABA06384/C
ID ABA06384 standard; cDNA; 1183 BP.
XX
XX ABA06384;
XX
XX 10-JAN-2002 (first entry)
XX
XX Human cDNA SEQ ID NO: 50.
XX
XX Human; gene therapy; neural disorder; immune system disorder;
XX muscular disorder; reproductive disorder; gastrointestinal disorder;
XX pulmonary disorder; cardiovascular disorder; renal disorder;
XX proliferative disorder; inflammation; ss.
OS Homo sapiens.
XX
XX WO200154474-A2.
XX
XX 02-AUG-2001.
XX
XX 17-JAN-2001; 2001WO-US01349.
XX
XX 31-JAN-2000; 2000US-179065P.
XX 04-FEB-2000; 2000US-180628P.
XX 24-FEB-2000; 2000US-184664P.
XX 02-MAR-2000; 2000US-186350P.
XX 16-MAR-2000; 2000US-189874P.
XX 17-MAR-2000; 2000US-190076P.
XX 18-APR-2000; 2000US-198123P.
XX 19-MAY-2000; 2000US-205515P.
XX 07-JUN-2000; 2000US-209467P.
XX 28-JUN-2000; 2000US-214886P.
XX 30-JUN-2000; 2000US-215135P.
XX 07-JUL-2000; 2000US-216647P.
XX 07-JUL-2000; 2000US-216880P.
XX 11-JUL-2000; 2000US-217487P.
XX 11-JUL-2000; 2000US-217496P.
XX 14-JUL-2000; 2000US-218290P.
XX 26-JUL-2000; 2000US-220963P.
XX 26-JUL-2000; 2000US-220964P.
XX 14-AUG-2000; 2000US-224518P.
XX 14-AUG-2000; 2000US-224519P.
XX 14-AUG-2000; 2000US-225213P.
XX 14-AUG-2000; 2000US-225214P.
XX 14-AUG-2000; 2000US-225266P.
XX 14-AUG-2000; 2000US-225267P.
XX 14-AUG-2000; 2000US-225268P.
XX 14-AUG-2000; 2000US-225270P.
XX 14-AUG-2000; 2000US-225447P.
XX 14-AUG-2000; 2000US-225757P.
XX 14-AUG-2000; 2000US-225758P.
XX 14-AUG-2000; 2000US-225759P.
XX 18-AUG-2000; 2000US-226279P.
XX 22-AUG-2000; 2000US-226681P.
XX 22-AUG-2000; 2000US-226686P.

PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
PI Rosen CA, Barash SC, Ruben SM;
XX WPI; 2001-451936/48.
DR P-PSDB; AAU18411.
XX
XX Isolated polypeptide for treating, preventing and/or prognosing
PT disorders of the endocrine system such as reproductive disorders,
PT endocrine cancers and also for testing and detection e.g. diagnosis -
XX
XX Claim 1; SEQ ID No 140; 604bp; English.
XX
XX Sequences AAS29511-AAS29736 represent cDNA molecules, which encode the
CC endocrine polypeptides of the invention. Endocrine polypeptides and their
CC associated polynucleotides of the invention are useful in the diagnosis,
CC treatment and prevention of various types of disorders in e.g. humans,
CC mice, rabbits, goats, horses, cats, dogs, chickens or sheep. A
CC pathological condition can be determined by determining the presence or
CC absence of a mutation in an endocrine polynucleotide. The treatable
CC disorders include autoimmune diseases such as rheumatoid arthritis,
CC hyperproliferative disorders such as neoplasms of the breast or liver,
CC cardiovascular disorders such as cardiac arrest, cerebrovascular
CC disorders such as cerebral ischemia, nervous system disorders such as
CC Alzheimer's disease, infections caused by bacteria, viruses and fungi,
CC ocular disorders such as corneal infection, endocrine disorders such as
CC premature labour and infertility, gastrointestinal disorders such as
CC Crohn's disease, renal disorders such as glomerulonephritis and
CC respiratory disorders such as asthma. The polypeptides can also be used
CC to aid wound healing, to prevent skin aging due to sunburn, to maintain
CC organs before transplantation, to regenerate tissues and in chemotaxis.
CC The polypeptides can also be used as a food additive or preservative to
CC increase or decrease storage capabilities.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO

Query Match 83.2%; Score 15.8; DB 22; Length 1183;
Best Local Similarity 89.5%; Pred. No. 1.4e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CTGGGGGCGATTATTCAC 19
DB 743 CTGGGGGCGATTATTCAC 725
IIIIIIII IIIII IIII
IIIIIIII IIIII IIII

RESULT 5
AAI63963/c
ID AAI63963 standard; cDNA; 1183 BP.
XX
XX AAI63963;
AC
XX
XX 22-OCT-2001 (first entry)
DT
XX
XX Human polynucleotide SEQ ID No 171.
DE
XX
XX Human; antiarthritic; antirheumatic; antiproliferative; vasotropic;
KW cerebroprotective; nootropic; neuroprotective; antibacterial; virucide;
KW fungicide; ophthalmological; cytostatic; immunosuppressive; nootropic;
KW neuroprotective; antiallergic; hepatotropic; antidiabetic;
KW antiinflammatory; anticancer; anticonvulsant; antibacterial;
KW antiparasitic; cardiac; gene therapy; cancer; immune disorder;
KW cardiovascular disorder; neurological disease; infection; human; ss.
XX
OS Homo sapiens.

XX
XX WO200155308-A2.
PN
XX
XX 02-AUG-2001.
PD
XX
XX 17-JAN-2001; 2001WO-US01309.
PE
XX
XX 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220863.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226779.
PR 22-AUG-2000; 2000US-0226881.
PR 22-AUG-2000; 2000US-0226882.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.

```
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239933.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246522.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251479.
PR 05-DEC-2000; 2000US-0251856.
PR 05-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
XX
PA (HUMA - ) HUMAN GENOME SCI INC.
XX

PI Rosen CA, Barash SC, Ruben SM;
XX WPI: 2001-488781/53.
DR P-PSDB: AAM43657.
XX
XX New isolated nucleic acids and polypeptides, useful for diagnosing,
PI treating and/or preventing human diseases and disorders -
XX
XX Claim 1; SEQ ID NO 171; 664pp + Sequence listing; English.
XX
XX The invention relates to human polynucleotides (AA163803-AA164012) and
CC the encoded proteins (AAM434497-AAM43660) useful for preventing, treating
CC or ameliorating medical conditions e.g. by protein or gene therapy. The
CC genes were isolated from a range of human tissues disclosed in the
CC specification. The nucleic acids, proteins, antibodies and (ant)agonists
CC are useful in the diagnosis, treatment and prevention of: (a) cancer,
CC e.g. breast and ovarian cancer and other cancers of the adrenal gland,
CC bone, bone marrow, breast, gastrointestinal tract, liver, lung, or
CC urogenital; (b) immune disorders e.g. Addison's disease, allergies,
CC autoimmune haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus,
CC Crohn's disease, multiple sclerosis, rheumatoid arthritis and ulcerative
CC colitis; (c) cardiovascular disorders such as myocardial ischaemias;
CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
CC and parasitic infections.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pcl_sequences.
XX
XX Sequence 1183 BP; 299 A; 272 C; 363 G; 247 T; 2 other;

Query Match 83.2%; Score 15.8; DB 22; Length 1183;
Best Local Similarity 89.5%; Pred. No. 1.4e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CTGGGGCGCATTATTCAC 19
DB 743 CTGGGGCGCATTATTCAC 725

RESULT 6
AAS31277/c
ID AAS31277 standard; cDNA: 1183 BP.
XX
AC AAS31277;
XX
DT 04-DEC-2001 (first entry)
XX
DE Human cDNA encoding a novel extracellular matrix protein, Seq ID No 91.
XX
XX Human; secreted extracellular matrix protein; ss; immunomodulatory;
XX Anti-HIV; antianemic; antirheumatic; antisclerotic; cardiac; vascular;
XX cerebroprotective; thrombolytic; antimicrobial; ophthalmic; cyostatic;
XX anti-12heimers; immune/autoimmune disease; HIV infection; anaemia;
XX human immunodeficiency virus; rheumatoid arthritis; multiple sclerosis;
XX cancer; hyperproliferative disorder; breast neoplasia; melanoma;
XX Searay syndrome; Gaucher's disease; neurological diseases;
XX Alzheimer's disease; Parkinson's disease; cardiovascular disorder;
XX cardiac arrest; tachycardia; angina; infection; corneal infections;
XX wound healing; immunogen; gene therapy; antisense; food additive.
XX
XX Homo sapiens.
XX OS
XX WO20015368-A1.
XX
XX 02-AUG-2001.
XX
XX 17-JAN-2001; 2001WO-US01348.
XX
XX 31-JAN-2000; 2000US-0179065.
XX 04-FEB-2000; 2000US-0180628.
XX 24-FEB-2000; 2000US-0184664.
XX 02-MAR-2000; 2000US-0186350.
```


CC The invention relates to isolated nucleic acid molecules encoding
CC novel human secreted extracellular matrix proteins (SPs). The
CC polynucleotides and proteins are used to prevent, treat a medical
CC condition in e.g. humans, mice, rabbits, goats, horses, cats, dogs,
CC chickens or sheep. For example, disorders associated with decreased
CC expression of SPs. The SP polynucleotide or a vector expressing them may
CC be administered to treat diseases by gene therapy. Antisense molecules
CC may be administered to down regulate expression of SPs by binding with
CC the cells own genes and preventing their expression. The polynucleotides
CC may also be used as DNA probes in diagnostic assays. The SPs may also be
CC used as antigens to produce antibodies and to identify modulators
CC (agonists and antagonists) of the SPs. The anti-(SP) antibodies and
CC antagonists may also be used to down regulate expression and activity of
CC SP and as diagnostic agents for detecting the presence of SPs in samples.
CC The disorders include for example: immune/autoimmune diseases (e.g. HIV
CC (human immunodeficiency virus) infections, anaemia, rheumatoid arthritis
CC and multiple sclerosis), cancers and hyperproliferative disorders (e.g.
CC melanomas, neoplasms of the breast or liver, Sezary syndrome and
CC Gaucher's disease), neurological diseases (e.g. Alzheimer's disease,
CC Parkinson's disease), cardio-/cerebrovascular disorders (e.g. cardiac
CC arrest, tachycardia and angina), infections caused by bacteria, viruses
CC and fungi and ocular disorders (e.g. corneal infections). Other uses
CC include wound healing, maintenance of organs before transplantation.

Query Match 83.2%; Score 15.8; DB 22; Length 1183;
Best Local Similarity 89.5%; Pred. No. 1.4e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 CTGGGGCAGTTATTCAC 19
||||| ||||| |||
DB 743 CTGGGGCGTATTTCAC 725

RESULT 7
AAS34773/c
ID AAS34773 standard; CDNA; 1183 BP.

AC AAS34773;

DT 04-DEC-2001 (first entry)

DE CDNA encoding novel human neoplastic disease associated polypeptide #7.

XX Human; neoplastic disease associated polypeptide; cancer; gene therapy;

KW hyperproliferative disorder; neural disorder; immune system disorder;

KW muscular disorder; reproductive disorder; gastrointestinal disorder;

KW pulmonary disorder; cardiovascular disorder; renal disorder;

KW neuroprotective; cytostatic; anti inflammatory; vasotropic; ss.

XX Homo sapiens.

OS WO200155163-A1.

PN 02-AUG-2001.

PD 17-JAN-2001; 2001WO-US01358.

XX 31-JAN-2000; 2000US-0179065.

PR 04-FEB-2000; 2000US-0180628.

PR 24-FEB-2000; 2000US-0184664.

PR 02-MAR-2000; 2000US-0186350.

PR 16-MAR-2000; 2000US-0189874.

PR 17-MAR-2000; 2000US-0190076.

PR 18-APR-2000; 2000US-0198123.

PR 19-MAY-2000; 2000US-0205515.

PR 07-JUN-2000; 2000US-0209467.

PR 28-JUN-2000; 2000US-0214886.

PR 30-JUN-2000; 2000US-0215135.

PR 07-JUL-2000; 2000US-0216647.

PR 07-JUL-2000; 2000US-0216880.

PR 11-JUL-2000; 2000US-0217487.

PR 11-JUL-2000; 2000US-0217496.

PR 14-JUL-2000; 2000US-0218290.

PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226881.
PR 22-AUG-2000; 2000US-0226886.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227182.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0232197.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237040.
PR 02-OCT-2000; 2000US-0237049.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.

PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246539.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
XX WPI; 2001-465558/50.
DR P-PSDB; AAU21574.
XX
XX
PT Novel polypeptides and polynucleotides useful as diagnostic reagents to
PT diagnose diseases or disorders associated with aberrant expression or
PT activity of polypeptides, and for treating cancers, rheumatoid
PT arthritis
XX
XX
PS Claim 4; SEQ ID NO 17; 687pp; English.
XX
XX
CC The present invention relates to the isolation of novel human neoplastic
CC disease associated polypeptides (AAU21568-AAU21851), and cDNA and DNA
CC sequences encoding for these polypeptides.
CC Invention are useful in the diagnosis, treatment, prevention and/or
CC prognosis of disorders involving neoplastic disease such as
CC hyperproliferative disorders (e.g. leukemia, bone cancer, bladder
CC cancer, brain stem glioma, adult liver cancer, childhood cerebellar
CC astrocytoma, or Hodgkin's lymphoma). The sequences of the invention may
CC also be useful for treating other disorders such as neural disorders,
CC immune system disorders, muscular disorders, reproductive disorders,
CC gastrointestinal disorders, pulmonary disorders, cardiovascular disorders

CC and renal disorders. The polynucleotide sequences of the invention are
CC also useful in gene therapy. AAU34767-AAU35050 represent cDNA sequences
CC encoding for the novel human neoplastic disease associated polypeptides
CC of the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 1183 BP; 299 A; 272 C; 363 G; 247 T; 2 other;
Query Match 83.2%; Score 15.8; DB 22; Length 1183;
Best Local Similarity 89.5%; Pred.No. 1.4e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 CTGGGGCGAGTTATTCAC 19
||||||| ||||| |||
Db 743 CTGGGGCGCTTATTCAC 725
RESULT 8
AB066601/c
ID AB066601 standard; cDNA; 1183 BP.
XX
AC AB066601;
XX
DT 23-AUG-2002 (first entry)
XX
DE Human polynucleotide SEQ ID NO 91.
XX
KW Human; nocrotic; neuroprotective; cytostatic; dermatological; virocidic;
KW immunosuppressive; antitumorigenic; anti-HIV; antibacterial; vulnary;
KW antiparkinsonian; antischistosomal; antileishmanial; cancer;
KW antihypertensive; hepatoprotective; cerebroprotective; antiinflammatory;
KW antiallergic; antidiabetic; antileukemic; anticonvulsant; antifungal;
KW antiparasitic; cardiac; immune disorder; cardiovascular disorder;
KW neurological disease; infection; nephrotropic; gene therapy; vaccine;
KW gene, ss.
XX
XX Homo sapiens.
OS
OS US2002042386-A1.
PN
XX
PD 11-APR-2002.
XX
XX
PF 17-JAN-2001; 2001US-0764870.
XX
XX 31-JAN-2000; 2000US-179065P.
PR 04-FEB-2000; 2000US-180628P.
PR 28-JUN-2000; 2000US-214886P.
PR 07-JUL-2000; 2000US-216647P.
PR 07-JUL-2000; 2000US-216880P.
PR 11-JUL-2000; 2000US-217487P.
PR 11-JUL-2000; 2000US-217496P.
PR 14-JUL-2000; 2000US-218290P.
PR 26-JUL-2000; 2000US-220963P.
PR 26-JUL-2000; 2000US-220964P.
PR 14-AUG-2000; 2000US-224518P.
PR 14-AUG-2000; 2000US-224519P.
PR 14-AUG-2000; 2000US-225267P.
PR 14-AUG-2000; 2000US-225268P.
PR 14-AUG-2000; 2000US-225270P.
PR 14-AUG-2000; 2000US-225447P.
PR 14-AUG-2000; 2000US-225757P.
PR 14-AUG-2000; 2000US-225758P.
PR 22-AUG-2000; 2000US-226868P.
PR 30-AUG-2000; 2000US-228924P.
PR 01-SEP-2000; 2000US-229287P.
PR 01-SEP-2000; 2000US-229343P.
PR 01-SEP-2000; 2000US-229344P.
PR 01-SEP-2000; 2000US-229345P.
PR 05-SEP-2000; 2000US-229509P.
PR 05-SEP-2000; 2000US-229513P.
PR 08-SEP-2000; 2000US-231413P.

PR 21-SEP-2000; 2000US-234223P.
PR 21-SEP-2000; 2000US-234274P.
PR 25-SEP-2000; 2000US-234997P.
PR 27-SEP-2000; 2000US-235834P.
PR 29-SEP-2000; 2000US-236327P.
PR 29-SEP-2000; 2000US-236367P.
PR 29-SEP-2000; 2000US-236368P.
PR 29-SEP-2000; 2000US-236369P.
PR 29-SEP-2000; 2000US-236370P.
PR 02-OCT-2000; 2000US-236802P.
PR 02-OCT-2000; 2000US-237037P.
PR 02-OCT-2000; 2000US-237038P.
PR 02-OCT-2000; 2000US-237039P.
PR 13-OCT-2000; 2000US-239335P.
PR 20-OCT-2000; 2000US-240960P.
PR 20-OCT-2000; 2000US-241785P.
PR 20-OCT-2000; 2000US-241809P.
PR 01-NOV-2000; 2000US-244617P.
PR 17-NOV-2000; 2000US-249299P.
PR 08-DEC-2000; 2000US-251856P.
PR 08-DEC-2000; 2000US-251868P.
PR 08-DEC-2000; 2000US-251869P.
XX
XX (ROSE/) ROSEN C A.
PA (RUBE/) RUBEN S M.
PA (BARA/) BARASH S C.
XX
PI Rosen CA, Ruben SM, Barash SC;
XX
DR WPI; 2002-470713/50.
DR P-PSDB; ABPA7926.
XX
PT New nucleic acid encoding human proteins, useful for diagnosis,
PT treatment and prevention of e.g. osteoporosis, also related
PT polypeptides and antibodies
XX
PS Claim 1; SEQ ID NO 91; 235pp + Sequence Listing; English.
XX
CC The invention relates to novel genes (ABQ66521-ABQ66785) and proteins
CC (ABQ47846-ABQ48110) useful for preventing, treating or ameliorating
CC medical conditions e.g. by protein or gene therapy. The genes are
CC isolated from a range of human tissues disclosed in the specification.
CC The nucleic acids, proteins, antibodies and (ant)agonists are useful
CC in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
CC and ovarian cancer and other cancers of the adrenal gland, bone, bone
CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune
CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative
CC colitis; (c) cardiovascular disorders such as myocardial ischaemias;
CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
CC and parasitic infections.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from USPTO at seqdata.uspto.gov/sequence.html?docid=999909764870.
XX
XX Sequence 1183 BP; 299 A; 272 C; 363 G; 247 T; 2 other:

Query Match 83.2%; Score 15.8; DB 24; Length 1183;
Best Local Similarity 89.5%; Pred. No. 1.4e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CTGGGGGAGTTATTGCAC 19
|||
Db 743 CTGGGGGCTGTATTCCAC 725

RESULT 9
AAZ52932/C
ID AAZ52932 standard; cDNA; 1490 BP.
XX

AC AAZ52932;
XX
DT 14-MAR-2000 (first entry)
XX
XX Human prostate tumor cDNA library derived EST fragment #75.
DE
XX
KW Pancreas; tumor; EST; expressed sequence tag; human; cytostatic;
KW treatment; ds.
XX
XX Homo sapiens.
OS
XX
XX DE19820190-A1.
PN
XX
PD 04-NOV-1999.
XX
PF 28-APR-1998; 98DE-1020190.
XX
XX 28-APR-1998; 98DE-1020190.
PR
XX
PA (META-) METAGEN GES GENOMFORSCHUNG MBH.
XX
PI Rosenthal A, Specht T, Hinzmann B, Schmitt A, Pilarsky C, Dahl E;
XX
XX WPI; 1999-621386/54.
DR P-PSDB; AAY74037, AAY74038, AAY74039.
XX
PT New human nucleic acid sequences from pancreatic tumors, and related
PT proteins
XX
PS Claim 2; Page 243; 502pp; German.
XX
XX
CC This invention describes novel polypeptides and their encoding nucleic
CC acids derived from human pancreatic tumor tissue which have cytostatic
CC activity. The sequences are also useful in producing pharmaceutical
CC compositions for treatment of pancreatic tumors. AAZ52858-253014
CC represent expressed sequence tag (EST) fragments derived from a human
CC pancreatic tumor cDNA library and which encode the proteins represented
CC in AAY73814-Y74252.
XX
XX Sequence 1490 BP; 319 A; 374 C; 464 G; 333 T; 0 other:

Query Match 83.2%; Score 15.8; DB 20; Length 1490;
Best Local Similarity 89.5%; Pred. No. 1.4e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CTGGGGGAGTTATTGCAC 19
|||
Db 1086 CTGGGGGCTGTATTCCAC 1068

RESULT 10
AAI63904/C
ID AAI63904 standard; cDNA; 2613 BP.
XX
XX AAI63904;
AC
XX
DT 22-OCT-2001 (first entry)
XX
XX Human polynucleotide SEQ ID NO 112.
DE
XX
XX Human; antiarthritic; antirheumatic; antiproliferative; vasotropic;
KW cerebroprotective; nootropic; neuroprotective; antibacterial; virucide;
KW fungicide; ophthalmological; cytostatic; immunosuppressive; nootropic;
KW neuroprotective; antiallergic; hepatotropic; antidiabetic;
KW antiinflammatory; antitumor; vulnery; anticonvulsant; antibacterial;
KW antiparasitic; cardiac; gene therapy; cancer; immune disorder;
KW cardiovascular disorder; neurological disease; infection; human; ss.
XX
XX Homo sapiens.
OS
XX
XX WO200155308-A2.
PN
XX
XX 02-AUG-2001.

XX 17-JAN-2001; 2001WO-US01309.
PF
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226686.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227109.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 08-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 12-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236368.

PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 29-SEP-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249219.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
(HUMA-) HUMAN GENOME SCI INC.
PA
XX
PI
XX
XX
DR
Rosen CA, Barash SC, Ruben SM;
MPI: 2001-488781/53.
P-PSDB; AAM43598.

XX New isolated nucleic acids and polypeptides, useful for diagnosing,
PT treating and/or preventing human diseases and disorders -
XX
XX Claim 1; SEQ ID NO 112; 664pp + Sequence Listing; English.
XX
CC The invention relates to human polynucleotides (AA163803-AA164012) and
CC the encoded proteins (AA434497-AA43660) useful for preventing, treating
CC or ameliorating medical conditions e.g. by protein or gene therapy. The
CC genes were isolated from a range of human tissues disclosed in the
CC specification. The nucleic acids, proteins, antibodies and (ant)agonists
CC are useful in the diagnosis, treatment and prevention of: (a) cancer,
CC e.g. breast and ovarian cancer and other cancers of the adrenal gland,
CC bone, bone marrow, breast, gastrointestinal tract, liver, lung, or
CC urogenital; (b) immune disorders e.g. Addison's disease, allergies,
CC autoimmune haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus,
CC Crohn's disease, multiple sclerosis, rheumatoid arthritis and ulcerative
CC colitis; (c) cardiovascular disorders such as myocardial ischaemias;
CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
CC and parasitic infections.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pft_sequences.
XX
SQ Sequence 2613 BP; 558 A; 670 C; 768 G; 614 T; 3 other;
Query Match 83.2%; Score 15.8; DB 22; Length 2613;
Best Local Similarity 89.5%; Pred. No. 1.5e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 CTGGGGGCGAGTATTGCAC 19
Db 2178 CTGGGGGCGTATTGCAC 2160
||||||| ||||| |||
RESULT 11
AAC77382/c
ID- AAC77382 standard; cDNA; 3463 BP.
XX
AC AAC77382;
XX
XX 08-FEB-2001 (first entry)
DE Human ORFX ORF2937 polynucleotide sequence SEQ ID NO:5873.
XX
XX Human: open reading frame; ORFX; detection; cytostatic; hepatotropic;
KW vulnery; antipsoptic; antiparkinsonian; nootropic; neuroprotective;
KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
KW hypotensive; dermatological; immunosuppressive; antiinflammatory;
KW antiviral; antibacterial; antifungal; antineumatic; antithyroid;
KW antihaemic; gene therapy; cancer; proliferative disorder; hypertension;
KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
KW cholesterol ester storage; systemic lupus erythematosus; infection;
KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
KW bone damage; cartilage damage; antiinflammatory disease; coagulation;
KW thrombosis; contraceptive; ss.
XX
XX Homo sapiens.
OS
XX
XX NO200058473-A2.
PN
XX
XX 05-OCT-2000.
PD
XX
XX 31-MAR-2000; 2000WO-US08621.
PF
XX
XX 31-MAR-1999; 99US-0127607.
PR
XX 02-APR-1999; 99US-0127636.
PR
XX 05-APR-1999; 99US-0127728.
PR
XX 30-MAR-2000; 2000US-0540763.

XX
PA (CURA-) CURAGEN CORP.
XX
XX Shimkels RA, Leach M;
PI
XX
XX WPI: 2000-602362/57.
DR
XX P-PSDB; ABA43173.
DR
XX
PT Novel nucleic acids and peptides derived from open reading frame X,
PT useful for treating e.g. cancers, proliferative disorders,
PT neurodegenerative disorders and cardiovascular disease -
XX
PS Claim 5; Page 5042-5044; 5507pp; English.
XX
CC AAC77446 to AAC77606 encode the proteins given in ABA40237 to ABA43397,
CC which represent the human ORFX open reading frames 1 to 3161. The ORFX
CC sequences have activities such as: cytostatic; hepatotropic; vulnery;
CC antipsoptic; antiparkinsonian; nootropic; neuroprotective;
CC osteopathic; anticonvulsant; antiarthritic; immunosuppressant;
CC immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;
CC antidiabetic; hypotensive; dermatological; immunosuppressive;
CC antiinflammatory; antibacterial; antiviral; antifungal; antineumatic;
CC antithyroid; and antihaemic. The sequences can be used for determining
CC the presence of or predisposition to, or preventing or treating
CC pathological conditions associated with an ORFX-associated disorder. The
CC nucleic acids can be used to express ORFX proteins in gene therapy.
CC
CC vectors. The proteins and nucleic acids may be used to treat cancers,
CC proliferative disorders, neurodegenerative disorders, osteoarthritis,
CC graft vs host disease, cardiovascular disease, diabetes mellitus,
CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,
CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
CC nocturnal haemoglobinuria, antiinflammatory disease; to enhance
CC coagulation; to inhibit thrombosis; and as a contraceptive.
XX
SQ Sequence 3463 BP; 721 A; 915 C; 1073 G; 753 T; 1 other;
Query Match 83.2%; Score 15.8; DB 21; Length 3463;
Best Local Similarity 89.5%; Pred. No. 1.5e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 CTGGGGGCGAGTATTGCAC 19
Db 2972 CTGGGGGCGTATTGCAC 2954
||||||| ||||| |||
RESULT 12
AAS68056
ID- AAS68056 standard; cDNA; 3531 BP.
XX
AC AAS68056;
XX
XX 13-FEB-2002 (first entry)
DE DNA encoding novel human diagnostic protein #3860.
XX
XX Human: chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
KW
XX
XX Homo sapiens.
OS
XX
XX WO200175067-A2.
PN
XX
XX 11-OCT-2001.
PD
XX
XX 30-MAR-2001; 2001WO-US08631.
PF
XX
XX 31-MAR-2000; 2000US-0540217.
PR
XX 23-AUG-2000; 2000US-0649167.
PR
XX (HYSE-) HYSEQ INC.
PA
XX


```

PR 18-JUL-2000; 2000US-219007P.
PR 13-DEC-2000; 2000US-255281P.
XX
XX
PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
XX PI Schlegel R, Endege WO, Monahan JE;
XX
XX DR WPI; 2001-662795/76.
XX
XX PT Novel isolated nucleic acid molecule associated with cancerous state of
XX prostate cells and correlating with presence of prostate cancer, useful
XX for detecting presence of prostate cancer, stage of prostate cancer
XX
XX PS Claim 1; Page 2090; 11750pp; English.
XX
XX CC The invention relates to an isolated nucleic acid molecule (I) comprising
XX a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
XX specification or its complement. (I) is useful for:
XX (a) assessing whether a patient is afflicted with prostate cancer;
XX (b) monitoring the progression of prostate cancer in a patient;
XX (c) assessing the efficacy of a test compound to inhibit prostate
XX cancer in a patient;
XX (d) assessing the efficacy of a therapy for inhibiting prostate cancer
XX in a patient;
XX (e) selecting a composition for inhibiting prostate cancer in a patient;
XX (f) assessing the prostate cell carcinogenic potential of a compound;
XX (g) determining whether prostate cancer has metastasized in a patient;
XX (h) assessing the aggressiveness or indolence of prostate cancer in a
XX patient;
XX (I) is also useful as a pharmacodynamic or pharmacogenomic marker.
XX
XX SQ Sequence 164 BP; 35 A; 40 C; 42 G; 47 T; 0 other;
XX
XX -Query Match 81.1%; Score 15.4; DB 23; Length 164;
XX Best Local Similarity 94.1%; Pred. No. 1.7e+02;
XX Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
XX
XX QY 1 CTGGGGGCGAGTTATTGC 17
XX ||||||| |||||||
XX DB 141 CTGGGGGCGATTATTGC 157
XX
RESULT 15
ABV3819
ID ABV3819 standard; cDNA; 427 BP.
XX
XX AC ABV3819;
XX
XX DT 16-SEP-2002 (first entry)
XX
XX DE Human prostate expression marker cDNA 33810.
XX
XX KW Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;
XX pharmacogenomic marker; gene; ss.
XX
XX OS Homo sapiens.
XX
XX PN WO200160860-A2.
XX
XX PD 23-AUG-2001.
XX
XX PF 20-FEB-2001; 2001WO-US05171.
XX
XX PR 17-FEB-2000; 2000US-183319P.
XX PR 16-MAR-2000; 2000US-189862P.
XX PR 25-MAY-2000; 2000US-207454P.
XX PR 09-JUN-2000; 2000US-211314P.
XX PR 18-JUL-2000; 2000US-219007P.
XX PR 13-DEC-2000; 2000US-255281P.
XX
XX PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
XX PI Schlegel R, Endege WO, Monahan JE;

```

```

XX
XX DR WPI; 2001-662795/76.
XX
XX PT Novel isolated nucleic acid molecule associated with cancerous state of
XX prostate cells and correlating with presence of prostate cancer, useful
XX for detecting presence of prostate cancer, stage of prostate cancer
XX
XX PS Claim 1; Page 7136; 11750pp; English.
XX
XX CC The invention relates to an isolated nucleic acid molecule (I) comprising
XX a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
XX specification or its complement. (I) is useful for:
XX (a) assessing whether a patient is afflicted with prostate cancer;
XX (b) monitoring the progression of prostate cancer in a patient;
XX (c) assessing the efficacy of a test compound to inhibit prostate
XX cancer in a patient;
XX (d) assessing the efficacy of a therapy for inhibiting prostate cancer
XX in a patient;
XX (e) selecting a composition for inhibiting prostate cancer in a patient;
XX (f) assessing the prostate cell carcinogenic potential of a compound;
XX (g) determining whether prostate cancer has metastasized in a patient;
XX (h) assessing the aggressiveness or indolence of prostate cancer in a
XX patient;
XX (I) is also useful as a pharmacodynamic or pharmacogenomic marker.
XX
XX SQ Sequence 427 BP; 86 A; 109 C; 115 G; 117 T; 0 other;
XX
XX -Query Match 81.1%; Score 15.4; DB 23; Length 427;
XX Best Local Similarity 94.1%; Pred. No. 1.9e+02;
XX Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
XX
XX QY 1 CTGGGGGCGAGTTATTGC 17
XX ||||||| |||||||
XX DB 174 CTGGGGGCGATTATTGC 190
XX
Search completed: July 1, 2003, 06:34:04
Job time : 106.231 secs

```


GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 1, 2003, 06:21:32 ; Search time 25.333 Seconds
(Without alignments)
230.008 Million cell updates/sec

Title: US-10-053-662A-30

Perfect score: 19

Sequence: 1 ctgggggcagttatgcac 19

Scoring table: IDENTITY_NUC

Gapop 10.0 ; Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued_Patents_NA.*

1: /cgn2-6/ptodata/1/ina/5A.COMB.seq.*

2: /cgn2-6/ptodata/1/ina/5B.COMB.seq.*

3: /cgn2-6/ptodata/1/ina/6A.COMB.seq.*

4: /cgn2-6/ptodata/1/ina/6B.COMB.seq.*

5: /cgn2-6/ptodata/1/ina/6C.COMB.seq.*

6: /cgn2-6/ptodata/1/ina/6D.COMB.seq.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysts of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	15.4	81.1	4316	1	US-08-317-450B-14
C 2	15.4	81.1	4316	3	US-08-800-593-14
C 3	15.4	81.1	5156	4	US-09-643-597-130
C 4	15.4	81.1	5200	1	US-08-317-450B-12
C 5	15.4	81.1	5200	3	US-08-800-593-12
C 6	15.4	81.1	11443	4	US-08-961-527-49
C 7	14.8	77.9	2085	4	US-09-352-990-9
C 8	14.8	77.9	5406	4	US-08-961-527-166
C 9	14.4	75.8	21	3	US-09-009-913-127
C 10	14.4	75.8	852	4	US-09-134-001C-2815
C 11	14.4	75.8	1080	4	US-09-134-001C-891
C 12	14.4	75.8	1958	4	US-08-665-034A-3
C 13	14.4	75.8	2200	2	US-08-626-685A-9
C 14	14.4	75.8	2200	4	US-08-993-088A-1
C 15	14.4	75.8	2200	4	US-08-993-424B-1
C 16	14.4	75.8	4403765	4	US-09-103-840A-2
C 17	14.4	75.8	4411529	4	US-09-103-840A-1
C 18	14.2	74.7	39	3	US-09-289-751-2
C 19	14.2	74.7	40	3	US-09-289-751-1
C 20	14.2	74.7	1494	4	US-09-071-035-341
C 21	14.2	73.7	3345	1	US-07-972-791-7
C 22	14.2	73.7	3361	1	US-07-972-791-6
C 23	13.8	72.6	139	4	US-09-141-027-11
C 24	13.8	72.6	465	4	US-09-134-001C-1373
C 25	13.8	72.6	1145	4	US-09-227-794-1
C 26	13.8	72.6	1380	2	US-08-748-947A-1
C 27	13.8	72.6	2520	3	US-08-968-563-10

28	13.8	72.6	2520	3	US-08-969-683A-10	Sequence 10, Appl
29	13.8	72.6	2520	4	US-09-297-928-6	Sequence 6, Appl
30	13.8	72.6	4104	4	US-08-961-527-169	Sequence 169, App
31	13.8	72.6	6012	1	US-08-375-709-14	Sequence 14, Appl
32	13.8	72.6	6012	1	US-08-752-929-14	Sequence 14, Appl
33	13.8	72.6	7147	4	US-08-961-527-23	Sequence 23, Appl
34	13.8	72.6	8898	4	US-08-961-527-69	Sequence 69, Appl
35	13.8	72.6	37895	1	US-08-375-709-1	Sequence 1, Appl
36	13.8	72.6	37895	1	US-08-752-929-1	Sequence 1, Appl
37	13.8	72.6	37895	3	US-09-090-793-12	Sequence 12, Appl
38	13.8	72.6	40138	3	US-09-090-793-12	Sequence 12, Appl
39	13.4	70.5	773	4	US-08-936-165A-163	Sequence 163, App
40	13.4	70.5	2234	4	US-08-993-088A-8	Sequence 8, Appl
41	13.4	70.5	2234	4	US-08-993-424B-8	Sequence 8, Appl
42	13.4	70.5	5631	4	US-09-052-469-1	Sequence 1, Appl
43	13.4	70.5	12912	2	US-08-460-751-1	Sequence 1, Appl
44	13.4	70.5	13807	4	US-09-052-469-5	Sequence 5, Appl
45	13.4	70.5	14060	3	US-08-658-136-4	Sequence 4, Appl

ALIGNMENTS

RESULT 1
US-08-317-450B-14/C
Sequence 14, Application US/08317450B
Patent No. 5660982
GENERAL INFORMATION:
APPLICANT: Tryggvason, Karl
APPLICANT: Kallunki, Pekka
APPLICANT: Pyke, Charles
TITLE OF INVENTION: Laminin Chains: Diagnostic and
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESSES:
ADDRESSEE: BANNER & ALLEGRETTI, LTD.
STREET: Ten South Wacker Drive
CITY: Chicago
STATE: Illinois
COUNTRY: USA
ZIP: 60606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/317,450B
FILING DATE: 04-OCT-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Chao, Mark
REGISTRATION NUMBER: 37,293
REFERENCE/DOCKET NUMBER: 94,778
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312-715-1000
TELEFAX: 312-715-1234
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 4316 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
FEATURE:
NAME/KEY: sig.peptide
LOCATION: 118..183
FEATURE:
NAME/KEY: CDS
LOCATION: 118..3453
FEATURE:
NAME/KEY: repeat_unit
LOCATION: 4021..4316

SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/317,450B
FILING DATE: 04-OCT-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Chao, Mark
REGISTRATION NUMBER: 37,293
REFERENCE/DOCKET NUMBER: 94,778
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312-715-1000
TELEFAX: 312-715-1234
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 5200 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
FEATURE:
NAME/KEY: sig_peptide
LOCATION: 118..183
FEATURE:
NAME/KEY: CDS
LOCATION: 118..3699
FEATURE:
NAME/KEY: polyA_site
LOCATION: 4433
FEATURE:
NAME/KEY: polyA_site
LOCATION: 5195
US-08-317-450B-12

Query Match 81.1%; Score 15.4; DB 1; Length 5200;
Best Local Similarity 94.1%; Pred. No. 33;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 3 GGGGGCAGTATTGAC 19
DB 1573 GAGGGCAGTATTGAC 1557

RESULT 5
US-08-800-593-12/c
Sequence 12, Application US/08800593
Patent No. 6143505
GENERAL INFORMATION:
APPLICANT: Tryggvason, Karl
APPLICANT: Kallunki, Pekka
APPLICANT: Pyke, Charles
TITLE OF INVENTION: Laminin Chains: Diagnostic and
TITLE OF INVENTION: Therapeutic Use
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: McDonnell Boenue Hulbert & Berghoff
STREET: 300 South Wacker Drive
CITY: Chicago
STATE: Illinois
COUNTRY: USA
ZIP: 60606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/800,593
FILING DATE: 18-FEB-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/317,450
FILING DATE: 04-OCT-1994
CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:
NAME: Chao, Mark
REGISTRATION NUMBER: 37,293
REFERENCE/DOCKET NUMBER: 94,778-B
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312-913-0001
TELEFAX: 312-913-0002
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 5200 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
FEATURE:
NAME/KEY: sig_peptide
LOCATION: 118..183
FEATURE:
NAME/KEY: CDS
LOCATION: 118..3699
FEATURE:
NAME/KEY: polyA_site
LOCATION: 4433
FEATURE:
NAME/KEY: polyA_site
LOCATION: 5195
US-08-800-593-12

Query Match 81.1%; Score 15.4; DB 3; Length 5200;
Best Local Similarity 94.1%; Pred. No. 33;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 3 GGGGGCAGTATTGAC 19
DB 1573 GAGGGCAGTATTGAC 1557

RESULT 6
US-08-961-527-49
Sequence 49, Application US/08961527
Patent No. 6420135
GENERAL INFORMATION:
APPLICANT: Charles Kunsch
TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences
NUMBER OF SEQUENCES: 391
CORRESPONDENCE ADDRESS:
ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/961,527
FILING DATE:
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Brookes, A. Anders
REGISTRATION NUMBER: 36,373
REFERENCE/DOCKET NUMBER: PB340P1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512
INFORMATION FOR SEQ ID NO: 49:
SEQUENCE CHARACTERISTICS:

LENGTH: 11443 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
US-08-961-527-49

Query Match 81.1%; Score 15.4; DB 4; Length 11443;
Best Local Similarity 94.1%; Pred. No. 38;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 TGGGGGCGAGTTATTGCA 18
|||||
Db 7079 TGGGGGCGAGTTATTGCA 7095

RESULT 7
US-09-352-990-9
Sequence 9, Application US/09352990
Patent No. 6235090
GENERAL INFORMATION:
APPLICANT: Famodu, Layo O.
APPLICANT: Orozco, Buddy
APPLICANT: Rafalski, Antoni
TITLE OF INVENTION: Plant Aminoacyl-tRNA Synthetase
FILE REFERENCE: BB-1191
CURRENT APPLICATION NUMBER: US/09/352,990
CURRENT FILING DATE: 1999-07-14
EARLIER APPLICATION NUMBER: 60/092,866
EARLIER FILING DATE: July 15, 1998
NUMBER OF SEQ ID NOS: 29
SOFTWARE: Microsoft Office 97
SEQ ID NO 9
LENGTH: 2085
TYPE: DNA
ORGANISM: Zea mays
US-09-352-990-9

Query Match 77.9%; Score 14.8; DB 4; Length 2085;
Best Local Similarity 88.9%; Pred. No. 60;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CTGGGGCGAGTTATTGCA 18
|||||
Db 1926 CTGGGGCGAGTTATTGCA 1943

RESULT 8
US-08-961-527-166/c
Sequence 166, Application US/08961527
Patent No. 6420135
GENERAL INFORMATION:
APPLICANT: Charles Kunsch
TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences
NUMBER OF SEQUENCES: 391
CORRESPONDENCE ADDRESS:
ADDRESSEE: Human genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4MB storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/961,527
FILING DATE:
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:

ATTORNEY/AGENT INFORMATION:
NAME: Brookes, A. Anders
REGISTRATION NUMBER: 36,373
REFERENCE/DOCKET NUMBER: PB340P1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512
INFORMATION FOR SEQ ID NO: 166:
SEQUENCE CHARACTERISTICS:
LENGTH: 5406 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
US-08-961-527-166

Query Match 77.9%; Score 14.8; DB 4; Length 5406;
Best Local Similarity 88.9%; Pred. No. 70;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 TGGGGGCGAGTTATTGCA 19
|||||
Db 4167 TGGGGGCGAGTTATTGCA 4150

RESULT 9
US-09-009-913-127
Sequence 127, Application US/09009913
Patent No. 6087485
GENERAL INFORMATION:
APPLICANT: Axy's Pharmaceuticals, Inc.
TITLE OF INVENTION: Asthma Related Genes
NUMBER OF SEQUENCES: 339
CORRESPONDENCE ADDRESS:
ADDRESSEE: Bozicevic & Reed, LLP
STREET: 285 Hamilton Ave, Suite 200
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94301
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/009,913
FILING DATE: 21-JAN-1998
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Sherwood, Pamela J
REGISTRATION NUMBER: 36,677
REFERENCE/DOCKET NUMBER: SEO-4P
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-327-3231
TELEFAX: 650-327-3231
TELEX:
INFORMATION FOR SEQ ID NO: 127:
SEQUENCE CHARACTERISTICS:
LENGTH: 21 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: Other
US-09-009-913-127

Query Match 75.8%; Score 14.4; DB 3; Length 21;
Best Local Similarity 83.3%; Pred. No. 45;
Matches 15; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 CTGGGGCGAGTTATTGCA 18

```

1      ZIP: 10036
2      COMPUTER READABLE FORM:
3      MEDIUM TYPE: Floppy disk
4      COMPUTER: IBM PC compatible
5      OPERATING SYSTEM: PC-DOS/MS-DOS
6      SOFTWARE: PatentIn Release #1.0, Version #1.30A
7      CURRENT APPLICATION DATA:
8      APPLICATION NUMBER: 05/08/626,685A

```

FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: White, John P.
REGISTRATION NUMBER: 28,678
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-278-0400
TELEFAX: 212-391-0525
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 2200 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA
FEATURE:
NAME/KEY: CDS
LOCATION: 46..414
FEATURE:
NAME/KEY: CDS
LOCATION: 1422..2171
US-08-626-685A-9

Query Match 75.8%; Score 14.4; DB 2; Length 2200;
Best Local Similarity 93.8%; Pred. No. 99;
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

1 CTGGGGGCGAGTTATG 16
|||||
Db - 1271 CTGGGGGCGAGTTATG 1256

RESULT 14
US-08-993-088A-1/c
Sequence 1, Application US/08993088A
Patent No. 6287855
GENERAL INFORMATION:
APPLICANT: Tan, Carina
APPLICANT: Sullivan, Kathleen
TITLE OF INVENTION: GALANIN RECEPTOR GALT2 AND
TITLE OF INVENTION: NUCLEOTIDES ENCODING SAME
NUMBER OF SEQUENCES: 20
CORRESPONDENCE ADDRESS:
ADDRESSEE: Merck & Co., Inc.
STREET: P.O. Box 2000, 126 E. Lincoln Ave.
CITY: Rahway
STATE: NJ
COUNTRY: USA
ZIP: 07065-0900
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: Windows
SOFTWARE: FASTSEQ for Windows Version 2.0b
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/993,088A
FILING DATE: 18-DEC-1997
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/033,851
FILING DATE: 27-DEC-1996
ATTORNEY/AGENT INFORMATION:
NAME: Heber, Sheldon O.
REGISTRATION NUMBER: 38,179
REFERENCE/DOCKET NUMBER: 19846
TELECOMMUNICATION INFORMATION:
TELEPHONE: 732-594-1958
TELEFAX: 732-594-4720
TELEX:
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2200 base pairs

TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
US-08-993-088A-1

Query Match 75.8%; Score 14.4; DB 4; Length 2200;
Best Local Similarity 93.8%; Pred. No. 99;
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

1 CTGGGGGCGAGTTATG 16
|||||
Db 1004 CTGGGGGCGAGTTATG 989

RESULT 15
US-08-993-424B-1/c
Sequence 1, Application US/08993424B
Patent No. 6337206

GENERAL INFORMATION:
APPLICANT: Tan, Carina
APPLICANT: Kolakowski, Lee F., Jr.
TITLE OF INVENTION: MOUSE GALANIN RECEPTOR GALT2 AND
TITLE OF INVENTION: NUCLEOTIDES ENCODING SAME
NUMBER OF SEQUENCES: 18
CORRESPONDENCE ADDRESS:
ADDRESSEE: Merck & Co., Inc.
STREET: P.O. Box 2000, 126 E. Lincoln Ave.
CITY: Rahway
STATE: NJ
COUNTRY: USA
ZIP: 07065-0900

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: Windows
SOFTWARE: FASTSEQ for Windows Version 2.0b
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/993,424B
FILING DATE: 18-DEC-1997
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/033,851
FILING DATE: 27-DEC-1996

ATTORNEY/AGENT INFORMATION:
NAME: Heber, Sheldon O.
REGISTRATION NUMBER: 38,179
REFERENCE/DOCKET NUMBER: 19846NP2
TELECOMMUNICATION INFORMATION:
TELEPHONE: 732-594-1958
TELEFAX: 732-594-4720
TELEX:

INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2200 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
US-08-993-424B-1

Query Match 75.8%; Score 14.4; DB 4; Length 2200;
Best Local Similarity 93.8%; Pred. No. 99;
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

1 CTGGGGGCGAGTTATG 16
|||||
Db 1004 CTGGGGGCGAGTTATG 989

Search completed: July 1, 2003, 07:24:00
Job time : 29.3333 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 1, 2003, 06:30:26 ; Search time 63.8205 Seconds

(Without alignments)
441.935 Million cell updates/sec

Title: US-10-053-662A-30

Perfect score: 19

Sequence: 1 ctgggggcagttatgcac 19

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1055720 seqs, 742224136 residues

Total number of hits satisfying chosen parameters: 2111440

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published_Applications_NA:*

- 1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
- 2: /cgn2_6/ptodata/1/pubpna/PC1_NEW_PUB.seq:*
- 3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
- 4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
- 5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
- 6: /cgn2_6/ptodata/1/pubpna/PC1S_PUBCOMB.seq:*
- 7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
- 8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
- 9: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
- 10: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
- 11: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
- 12: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
- 13: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
- 14: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	15.8	83.2	1183	9	US-10-125-540-91
2	15.8	83.2	1183	9	US-10-103-313-17
3	15.8	83.2	1183	10	US-09-764-870-91
4	15.8	83.2	1183	10	US-09-764-853-50
5	15.4	81.1	129	10	US-09-998-598-1920
6	15.4	81.1	4316	9	US-10-227-738-14
7	15.4	81.1	4316	10	US-09-756-071B-12
8	15.4	81.1	5156	10	US-09-735-705-130
9	15.4	81.1	5156	10	US-09-850-716A-130
10	15.4	81.1	5156	10	US-09-897-778-130
11	15.4	81.1	5175	9	US-10-171-311-114
12	15.4	81.1	5200	10	US-09-756-071B-12
13	15.4	81.1	5460	12	US-10-044-090-558
14	15.4	81.1	9542	9	US-10-092-154-1809
15	15.4	81.1	9542	10	US-09-764-847-1809
16	15.4	81.1	9542	10	US-09-764-847-1809
17	14.8	77.9	363	10	US-09-783-590-8880
18	14.8	77.9	471	10	US-09-864-761-16656
19	14.8	77.9	475	9	US-09-923-844B-3

20	14.8	77.9	566	10	US-09-864-761-9134	Sequence 9134, Ap
21	14.8	77.9	694	10	US-09-764-877-3567	Sequence 3567, Ap
22	14.8	77.9	694	10	US-09-764-877-3569	Sequence 3569, Ap
23	14.8	77.9	867	9	US-09-764-787-360	Sequence 360, App
24	14.8	77.9	3038	9	US-10-156-610-83	Sequence 83, App1
25	14.8	77.9	14155	9	US-10-108-605-102	Sequence 102, App1
26	14.4	75.8	264	9	US-10-083-357-232	Sequence 232, App
27	14.4	75.8	339	10	US-09-746-284-40	Sequence 40, App1
28	14.4	75.8	392	9	US-09-918-995-7055	Sequence 7055, Ap
29	14.4	75.8	446	9	US-09-918-995-26726	Sequence 26726, A
30	14.4	75.8	464	9	US-09-918-995-30742	Sequence 30742, A
31	14.4	75.8	487	9	US-09-918-995-29135	Sequence 29135, A
32	14.4	75.8	528	10	US-09-746-284-33	Sequence 33, App1
33	14.4	75.8	2200	7	US-08-899-112-9	Sequence 9, App1
34	14.4	75.8	3309400	9	US-09-738-626-1	Sequence 1, App1
35	14.2	74.7	401	10	US-09-960-352-600	Sequence 600, App
36	14.2	74.7	410	10	US-09-880-107-285	Sequence 285, App
37	14.2	74.7	412	10	US-09-864-761-3627	Sequence 3627, Ap
38	14.2	74.7	472	9	US-10-198-846-517	Sequence 517, App
39	14.2	74.7	843	9	US-10-073-979-14	Sequence 14, App1
40	14.2	74.7	1008	10	US-09-974-300-2753	Sequence 2753, Ap
41	14.2	74.7	1154	10	US-09-833-381-71	Sequence 71, App1
42	14.2	74.7	1386	9	US-09-738-626-1620	Sequence 1620, App
43	14.2	74.7	1498	9	US-09-808-602-26	Sequence 26, App1
44	14.2	74.7	1498	9	US-10-073-979-11	Sequence 11, App1
45	14.2	74.7	1498	9	US-09-800-198-24	Sequence 24, App1

ALIGNMENTS

RESULT 1
US-10-125-540-91/c
; Sequence 91, Application US/10125540
; Publication No. US20030059875A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PTZ1AC1
; CURRENT APPLICATION NUMBER: US/10/125,540
; CURRENT FILING DATE: 2002-04-19
; PRIOR APPLICATION removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 646
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 91
; LENGTH: 1183
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (72)
; OTHER INFORMATION: n equals a,t,g, or c
; US-10-125-540-91

Query Match 83.2%; Score 15.8; DB 9; Length 1183;
Best Local Similarity 89.5%; Pred. No. 77;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CTGGGGCAGTATTGCAC 19
||||| ||||| |||
Db 743 CTGGGGCAGTATTGCAC 725

RESULT 2
US-10-103-313-17/c
; Sequence 17, Application US/10103313
; Publication No. US20030082758A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PTZ07C1
; CURRENT APPLICATION NUMBER: US/10/103,313
; CURRENT FILING DATE: 2002-03-12

```

; NUMBER OF SEQ ID NOS: 653
; Prior Application removed - See File Wrapper or Palm
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 17
; LENGTH: 1183
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (72)
; OTHER INFORMATION: n equals a,t,g, or c
US-10-103-313-17

Query Match
Best Local Similarity 83.2%; Score 15.8; DB 9; Length 1183;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CTGGGGCAGTTATTGCAC 19
   ||||| ||||| |||
Db 743 CTGGGGCTGTATTCCAC 725

RESULT 3
US-09-764-870-91/c
; Sequence 91, Application US/09764870
; Patent No. US20020042386A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT214
; CURRENT APPLICATION NUMBER: US/09/764,870
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 646
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 91
; LENGTH: 1183
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (72)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-764-870-91

Query Match
Best Local Similarity 83.2%; Score 15.8; DB 10; Length 1183;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CTGGGGCAGTTATTGCAC 19
   ||||| ||||| |||
Db 743 CTGGGGCTGTATTCCAC 725

RESULT 4
US-09-764-853-50/c
; Sequence 50, Application US/09764853
; Patent No. US20020090672A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PJ206
; CURRENT APPLICATION NUMBER: US/09/764,853
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 939
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 50
; LENGTH: 1183
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
```

```

; LOCATION: (72)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-764-853-50

Query Match
Best Local Similarity 83.2%; Score 15.8; DB 10; Length 1183;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CTGGGGCAGTTATTGCAC 19
   ||||| ||||| |||
Db 743 CTGGGGCTGTATTCCAC 725

RESULT 5
US-09-998-598-1920/c
; Sequence 1920, Application US/09998598
; Patent No. US20020150922A1
; GENERAL INFORMATION:
; APPLICANT: Stolk, John A.
; APPLICANT: Xu, Jiangchun
; APPLICANT: Chenault, Ruth A.
; APPLICANT: Meagher, Madeline Joy
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE REFERENCE: 210121.561
; CURRENT APPLICATION NUMBER: US/09/998,598
; CURRENT FILING DATE: 2001-11-16
; NUMBER OF SEQ ID NOS: 2606
; SOFTWARE: Corlax Invention Disclosure Database
; SEQ ID NO 1920
; LENGTH: 129
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-998-598-1920

Query Match
Best Local Similarity 81.1%; Score 15.4; DB 10; Length 129;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 GGGGCGCATTATTGCAC 19
   ||||| ||||| |||
Db 49 GAGGCGCATTATTGCAC 33

RESULT 6
US-10-227-738-14/c
; Sequence 14, Application US/10227738
; Publication No. US20030100529A1
; GENERAL INFORMATION:
; APPLICANT: Trygvaason, Karl
; APPLICANT: Kallunki, Pekka
; APPLICANT: Pyke, Charles
; TITLE OF INVENTION: Laminin Chains: Diagnostic and
; THERAPEUTIC USE
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: McDonnell Boehnen Hulbert & Berghoff
; STREET: 300 South Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/227,738
; FILING DATE: 26-Aug-2002
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/800,593
```


Qy		3	GGGGGAGTTATTGCAC	19
Db		1521	GAGGGCAGTTATTGCAC	1505

```

RESULT 9
US-09-850-716A-130/c
: Sequence 130, Application US/09850716A
: Patent No. US2002011539A1
: GENERAL INFORMATION:
:
: APPLICANT: Kalos, Michael D.
: APPLICANT: McNeill, Patricia D.
:
: TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
: OF LUNG CANCER
:
: FILE REFERENCE: 210121.455C15
:
: CURRENT APPLICATION NUMBER: US/09/850,716A
:
: CURRENT FILING DATE: 2001-05-07
:
: NUMBER OF SEQ ID NOS: 440
:
: SOFTWARE: FastSeq for Windows Version 3.0
:
: SEQ ID NO 130
:
: LENGTH: 5156
:
: TYPE: DNA
:
: ORGANISM: Homo sapien
:
: US-09-850-716A-130

```

Query Match	81.1%	Score 15.4;	DB 10;	Length 5156;
Best Local Similarity	94.1%;	Pred. No. 1.5e+02;		
Matches 16; Conservative	0;	Mismatches 1;	Indels 0;	Gaps 0;

```

QY      3 GGGGGCAGTTATTGCAC 19
          | | | | | | | | | |
Db     1521 GAGGGCAGTTATTGCAC 1505

```

```

RESULT 10
US-09-897-778-130/c
Sequence 130, Application US/09897778
Patent No. US20020147143A1
GENERAL INFORMATION:
APPLICANT: Wang, Tongtong
APPLICANT: Marnierakis, Margarita
APPLICANT: Fanger, Gary R.
APPLICANT: Vedvick, Thomas S.
APPLICANT: Carter, Darick
APPLICANT: Watanabe, Yoshinhiro
APPLICANT: Henderson, Robert A.
APPLICANT: Peckham, David W.
APPLICANT: Fanger, Neil
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
TITLE OF INVENTION: AND DIAGNOSIS OF LUNG CANCER
FILE REFERENCE: 210121.455C16
CURRENT APPLICATION NUMBER: US/09/897,778
CURRENT FILING DATE: 2001-06-28
NUMBER OF SEQ ID NOS: 467
SOFTWARE: FASTSEQ for Windows Version 4.0
SEQ ID NO 130
LENGTH: 5156
TYPE: DNA
ORGANISM: Homo sapiens
US-09-897-778-130

```

Query Match	81.1%;	Score 15.4;	DB 10;	Length 5156;
Best Local Similarity	94.1%;	Pred. No. 1.5e+02;		
Matches 16; Conservative	0;	Mismatches 1;	Indels 0;	Gaps 0;

```
QY      3 GGGGCGAGTTATTCAC 19
          | | | | | | | | | |
Db     1521 GAGGCGAGTTATTCAC 1505
```

RESULT 11
US-10-171-311-114/c

Sequence 114 Application US/10171311
Publication No. US20030087270A1
GENERAL INFORMATION:
APPLICANT: Schlegel, Robert
APPLICANT: Chen, Yan
APPLICANT: Zhao, Xumei
APPLICANT: Monahan, John
APPLICANT: KamalKar, Shubhangi
APPLICANT: Glat, Karen
APPLICANT: Gannavarapu, Manjula
APPLICANT: Hoersh, Sebastian
TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY
TITLE OF INVENTION: OF CERVICAL CANCER
FILE REFERENCE: MRI-035
CURRENT APPLICATION NUMBER: US/10/171,311
CURRENT FILING DATE: 2002-06-12
PRIOR APPLICATION NUMBER: US 60/298,159
PRIOR FILING DATE: 2001-06-13
PRIOR APPLICATION NUMBER: US 60/298,155
PRIOR FILING DATE: 2001-06-13
PRIOR APPLICATION NUMBER: US 60/335,936
PRIOR FILING DATE: 2001-11-14
NUMBER OF SEQ ID NOS: 238
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 114
LENGTH: 5175
TYPE: DNA
ORGANISM: Homo sapiens
US-10-171-311-114

Query Match	81.1%;	Score 15.4;	DB 9;	Length 5175;
Best Local Similarity	94.1%;	Pred. No. 1.5e+02;		
Matches 16;	Conservative	0;	Mismatches 1;	Indels 0;
				Gaps 0;

```

QY      3 GGGGCGATTATGCAC 19
          | | | | | | | | | |
Db     1545 GAGGCGAGTTATGCAC 1529

```

RESULT 12
 US-10-227-738-12/c
 Sequence 12, Application US/10227738
 Publication No. US20030100529A1
 GENERAL INFORMATION:
 APPLICANT: Tryggyason, Karl
 Kallunki, Pekka
 Pyke, Charles
 TITLE OF INVENTION: Laminin Chains: Diagnostic and
 Therapeutic Use
 NUMBER OF SEQUENCES: 19
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: McDonnell Boehnen Hulbert & Berghoff
 STREET: 300 South Wacker Drive
 CITY: Chicago
 STATE: Illinois
 COUNTRY: USA
 ZIP: 60606
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/10/227,738
 FILING DATE: 26-Aug-2002
 CLASSIFICATION: 435
 PRIORITY APPLICATION DATA:
 APPLICATION NUMBER: US/08/800,593
 FILING DATE: 18-FEB-1997
 APPLICATION NUMBER: US 08/317,450
 FILING DATE: 04-OCT-1994
 ATTORNEY/AGENT INFORMATION:

; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1809
; LENGTH: 9542
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-092-134-1809

Query Match 81.1%; Score 15.4; DB 9; Length 9542;
Best Local Similarity 94.1%; Pred. No. 1.6e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 TGGGGGCAGTTATGCA 18
|||||
DB 1525 TGGGGGCAGTTATGCA 1541

Search completed: July 1, 2003, 07:26:19
Job time : 65.8205 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 1, 2003, 03:57:39 ; Search time 859.385 Seconds
(without alignments)
358.063 Million cell updates/sec

Title: US-10-053-662A-30

Perfect score: 19

Sequence: 1 ctgggggcagctatgcac 19

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

EST:*
1: em_estb1:*
2: em_estb2:*
3: em_estb3:*
4: em_estb4:*
5: em_estb5:*
6: em_estb6:*
7: em_estb7:*
8: em_estb8:*
9: em_estb9:*
10: em_estb10:*
11: em_estb11:*
12: em_estb12:*
13: em_estb13:*
14: em_estb14:*
15: em_estb15:*
16: em_estb16:*
17: em_estb17:*
18: em_estb18:*
19: em_estb19:*
20: em_estb20:*
21: em_estb21:*
22: em_estb22:*
23: em_estb23:*
24: em_estb24:*
25: em_estb25:*
26: em_estb26:*
27: em_estb27:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	17.4	91.6	1201	13	BM046418
2	16.4	86.3	144	9	AL842103
3	16.4	86.3	183	9	AL194150
4	16.4	86.3	271	14	M78847
5	16.4	86.3	396	12	BF850108
6	16.4	86.3	457	10	AW134919

Result No.	Score	Query Match	Length	DB ID	Description
7	16.4	86.3	458	13	BM046418
8	16.4	86.3	489	13	AL048818
9	16.4	86.3	524	13	BF076955
10	16.4	86.3	544	10	AM655709
11	16.4	86.3	701	17	BH428192
12	16.4	86.3	701	17	AG132023
13	16.4	86.3	716	13	BU153383
14	16.4	86.3	752	12	BG282508
15	16.4	86.3	866	17	BH442712
16	16.4	86.3	984	17	BG282706
17	16.4	86.3	1101	17	CNS05M09
18	16.4	86.3	1298	14	BQ049927
19	16.4	86.3	1469	16	BE130964
20	16.4	86.3	1682	17	AG056551
21	16.4	86.3	188	12	BE857007
22	16.4	86.3	297	9	AI858860
23	16.4	86.3	297	14	T36249
24	16.4	86.3	301	12	BF916711
25	16.4	86.3	352	9	AU110219
26	16.4	86.3	364	13	BI033206
27	16.4	86.3	370	10	AM801445
28	16.4	86.3	371	10	AM801444
29	16.4	86.3	374	13	BI161676
30	16.4	86.3	406	12	BF482335
31	16.4	86.3	412	10	AM778993
32	16.4	86.3	414	10	AM514348
33	16.4	86.3	426	9	AA742413
34	16.4	86.3	435	10	AM801448
35	16.4	86.3	446	9	AA398958
36	16.4	86.3	451	9	AI924185
37	16.4	86.3	461	9	AA93540
38	16.4	86.3	469	17	AQ278480
39	16.4	86.3	470	9	AI337830
40	16.4	86.3	487	9	AA703508
41	16.4	86.3	487	12	BE748722
42	16.4	86.3	489	14	BQ319898
43	16.4	86.3	490	12	BF889999
44	16.4	86.3	492	9	AA459604
45	16.4	86.3	492	12	BF890005

ALIGNMENTS

RESULT 1
LOCUS BM046418 1201 bp mRNA linear EST 07-NOV-2001
DEFINITION 603626447F1 NIH_MGC_40 Homo sapiens cDNA clone IMAGE:5453002 5',
ACCESSION BM046418
VERSION BM046418.1 GI:1675685
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 1201)
AUTHORS NIH-MGC <http://mgi.nci.nih.gov/>
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cga@bbs-rmail.nih.gov
Tissue Procurement: DCTD/DDP
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LHC1945 row: n column: 11
High quality sequence start: 3
High quality sequence stop: 117.
Location/Qualifiers

source
1. .1201
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5453002"
/clone.lib="NIH_MGC_40"
/tissue_type="carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: prostate; Vector: pOTB7; Site:1; XhoI;
Site_2: EcoRI; cDNA made by oligo-dT priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCACGAG(G). Library constructed by
Ling Hong in the laboratory of Gerald M. Rubin (University
of California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies).
Note: this is a NIH_MGC library."

BASE COUNT 358 a 416 c 239 g 187 t 1 others

ORIGIN

Query Match 91.6%; Score 17.4; DB 13; Length 1201;
Best Local Similarity 94.7%; Pred. No. 1.9e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CTGGGGCAGTTATTCAC 19
|||||

DB 700 CTGGGGCAGTTATTCAC 718

RESULT 2
AL842103 144 bp mRNA linear EST 30-JUL-2002
LOCUS AL842103 F000F Takifugu rubripes cDNA clone F000F15a1, mRNA
DEFINITION sequence.
ACCESSION AL842103
VERSION AL842103.1 GI:22019931
KEYWORDS EST.
SOURCE Takifugu rubripes.
ORGANISM Takifugu rubripes.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
Tetraodontidae; Takifugu.
1 (bases 1 to 144)
Clark, M.S.
Takifugu rubripes ESTs
Unpublished (2002)
Contact: Clark MS
MRC Human Genome Mapping Project Resource Centre
Hinxton, Cambridge, CB10 1SB, UK
Email: biohelp@hmp.mrc.ac.uk
Vector: pME18S-FL3
V-type: phagemid
PRIMER: ME-735FW
Library created by Koichi Kawakami, Masahide Sasaki, Yutaka Suzuki,
Sumio Sugano
The Institute of Medical Science, The University of Tokyo,
Shirokanedai, Minato-ku, Tokyo 108-8639, Japan
and
Kiyoshi Kikuchi, Shugo Watabe
Laboratory of Aquatic Molecular Biology and Biotechnology, Graduate
School of Agricultural and Life Sciences, The University of Tokyo,
Bunkyo-ku, Tokyo 108-8639, Japan
Library sequenced by Melody S. Clark and Amanda Thompson MRC Human
Genome Mapping Project Resource Centre Hinxton, Cambridge, CB10
1SB, UK.

FEATURES
source
1. .144
/organism="Takifugu rubripes"
/db_xref="taxon:31033"
/clone="F000F15a1"
/clone.lib="F000F"
/sex="female"
/tissue_type="fin"
/dev_stage="adult"

BASE COUNT 37 a 40 c 38 g 29 t
ORIGIN

Query Match 86.3%; Score 16.4; DB 9; Length 144;
Best Local Similarity 94.4%; Pred. No. 4.4e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CTGGGGCAGTTATTCAC 18
|||||

DB 116 CTGGGGCAGTTATTCAC 133

RESULT 3
A1194150 183 bp mRNA linear EST 13-OCT-1998
LOCUS ue75h02.r1 Soares_NMPu Mus musculus cDNA clone IMAGE:1496979 5'
DEFINITION similar to gb:X71642 M.musculus GEC-154 mRNA (MOUSE);, mRNA
sequence.
ACCESSION A1194150
VERSION A1194150.1 GI:3745357
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 183)
Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,
Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,
Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,
Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and
Waterston, R.
The WashU-HMI Mouse EST Project
Unpublished (1996)
Contact: Marra M/Mouse EST Project
WashU-HMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@wustl.wustl.edu
This clone is available royalty-free through LML; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:934583
Trace considered overall poor quality
Seq primer: -28m13 rev2 ET from Amersham
High quality sequence stop: 1.

FEATURES
source
1. .183
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="IMAGE:1496979"
/clone.lib="Soares_NMPu"
/sex="female"
/dev_stage="adult"
/lab_host="DH10B"
/note="Organ: uterus; Vector: pT73D-Pac (Pharmacia) with
a modified polylinker; 1st strand cDNA was prepared from
pregnant mouse uterus, and was then primed with a Not I -
oligo(dT) primer. Double-stranded cDNA was ligated to Eco
RI adaptors (Pharmacia), digested with Not I and cloned
into the Not I and Eco RI sites of the modified pT73
vector. Library is normalized. Library was constructed by
Bento Soares and M. Fatima Bernaldo."

BASE COUNT 74 a 26 c 48 g 35 t
ORIGIN

Query Match 86.3%; Score 16.4; DB 9; Length 183;
Best Local Similarity 94.4%; Pred. No. 4.6e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 TGGGGCAGTTATTCAC 19
|||||

Db		58	TGGTGCAGCTATTGCAC	75	
RESULT 4					
LOCUS	M78847				
DEFINITION	M78847	271 bp	mRNA	linear	EST 26-MAY-1992
VERSION	E5700995	Hippocampus, Striatum	cDNA	(cat. #936205)	Homo sapiens cDNA
KEYWORDS	clone HHCmH22,	mRNA sequence.			
SOURCE	M78847.1	GI:273162			
ORGANISM	EST.				
	human.				
	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.				
	1 (bases 1 to 271)				
TITLE	Adams,M.D., Dubnick,M., Kerlavage,A.R., Moreno,R., Kelley,J.M.,				
JOURNAL	Utterback,T.R., Nagle,J.W., Fields,C. and Venter,J.C.				
MEDLINE	Sequence identification of 2,375 human brain genes				
COMMENT	Nature 355, 632-634 (1992)				
	92168112				
	Contact: Kerlavage, AR				
	Bioinformatics				
	The Institute for Genomic Research				
	9712 Medical Center Drive, Rockville, MD 20850 USA				
	Tel.: 3018699056				
	Fax: 3018699423				
	Email: arkerlav@tigr.org				
	Seq primer: M13 Forward.				
FEATURES	Location/Qualifiers				
SOURCE	1..271				
	/organism="Homo sapiens"				
	/db_xref="ATCC (lnhost):78301"				
	/db_xref="GDB:D051828E"				
	/db_xref="taxon:9606"				
	/clone="HHCmH22"				
	/clone_lib="Hippocampus, Striatum (cat. #936205)"				
	/note=Vector: lambdaZAP-II; female, 2 years; oligo-dT +				
	random primed cDNA synthesis; lambdaZAP-II vector, 1.0kb				
	average insert size."				
BASE COUNT	54 a	98 c	57 g	60 t	2 others
ORIGIN					
Query Match		86.3%	Score 16.4:	DB 14:	Length 271:
Best Local Similarity		89.5%	Pred. No. 4.9e+02:		
Matches	17: Conservative	0: Mismatches	2: Indels	0: Gaps	0:
Oy	1 CTGGGGCGACTATTGCAC	19			
Db					
	218 CTGGGGCGCTGTATTNCAC	236			
RESULT 5					
LOCUS	BF850108	396 bp	mRNA	linear	EST 16-JAN-2001
DEFINITION	KCQ-EN0080-161100-021-a05 EN0080	Homo sapiens cDNA,			mRNA sequence.
ACCESSION	BF850108				
VERSION	BF850108.1	GI:12237258			
KEYWORDS	EST.				
SOURCE	human.				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.				
	1 (bases 1 to 396)				
	Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,				
	Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,				
	Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bala,G.S., Simpson,D.H.,				
	Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare				
	,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and				
	Simpson,A.J.				
	Shotgun sequencing of the human transcriptome with ORF expressed				
	sequence tags				
TITLE	Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)				
JOURNAL					

```

MEDLINE
COMMENT      20202663
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=RC06t2-RC0-EN0080-16100-021-a056t3-2000-11-16&tt4-1)
Seq primer: puc 18 forward
High quality sequence start: 63
High quality sequence stop: 396.
location/Qualifiers
1. 396
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_id="EN0080"
/dev_stage="Adult"
/note="Organ: lung,normal; Vector: puc18; Site_1: SmaI;
Site_2: SmaI; A mini-library was made by cloning products
derived from ORESTS PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the pUC 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."
BASE COUNT      95 a 90 c 110 g 101 t
ORIGIN
Query Match      86.3%; Score 16.4; DB 12; Length 396;
Best Local Similarity 94.4%; Pred. No. 5.2e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY      1 CTGGGGCAGCTATTGCA 18
11111111111111111111
Db      247 CTGGGGCAGCTATTGCA 264
RESULT 6
AM134919/c
LOCUS      AM134919      457 bp      mRNA      linear      EST 28-OCT-1999
DEFINITION      U1-H-B11-abr-d-05-0-UI.s1 NCI-CGAP-Sub3 Homo sapiens cDNA clone
IMAGE:2712633 3', mRNA sequence.
ACCESSION      AM134919
VERSION
KEYWORDS
SOURCE      AM134919.1 GI:6138465
EST.
human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 457)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
The sequence contained an oligo-dT track that was present in the
oligonucleotide that was used to prime the synthesis of first
strand cDNA and therefore this may represent a bonafide poly A
tail. cDNA library Preparation: M.B. Soares lab clone distribution:
NCI-CGAP clone distribution information can be found through the
I.M.A.G.E. Consortium/ILN at:
www.bio.lnl.gov/dbdp/Image/Image.html
Seq primer: M13 Forward
POLYA=yes.
location/Qualifiers
1. 457
/organism="Homo sapiens"
/db_xref="taxon:9606"

```

[illegible]

COMMENT	Contact: Simpson A.J.G. Laboratory of Cancer Genetics Ludwig Institute for Cancer Research Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil Tel: +55-11-2704922 Fax: +55-11-2707001 Email: asimpson@ludwig.org.br This sequence was derived from the FAPESP/LICR Human Cancer Genome project. This entry can be seen in the following URL (http://www.ludwig.org.br/scripts/gethtml2.pl?l1=PM0&t2=PM0-ht1311-270401-002-B09&t3=2001-04-27&t4=1) Seq primer: puc 18 forward High quality sequence start: 14 High quality sequence stop: 457. Location/Qualifiers			
FEATURES	1..458			
SOURCE	/organism="Homo sapiens" /db_xref="taxon:9606" /clone_lib="HT1311" /dev_stage="Adult" /note="Organ: head,neck; Vector: puc18; Site:1: Smal; Site:2: SmaI; A mini-library was made by cloning products derived from ORSTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."			
BASE COUNT	122 a	123 c	100 g	113 t
ORIGIN				
Query Match	86.3%	Score 16.4:	DB 13:	Length 458;
Best Local Similarity	94.4%;	Pred. No. 5.4e+02;		
Matches	17; Conservative	0; Mismatches	1; Indels	0; Gaps 0;
Oy	1 CTGGGGCGAGTATTGCA 18			
Db	254 CTGGGGCGACTATTGCA 237			
RESULT 8	489 bp mRNA linear EST 29-FEB-2000			
AL048818/c				
LOCUS	DKFZp434F2018_r1 434 (synonym: htes3) Homo sapiens cDNA clone			
DEFINITION	DKFZp434F2018, mRNA sequence.			
ACCESSION	AL048818			
VERSION	AL048818.1 GI:4728127			
KEYWORDS	EST.			
SOURCE	human.			
ORGANISM	Homo sapiens			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 489) Ottewaelder,B., Obermaier,B., Mewes,H.W., Gaassenhuber,J. and Wiemann,S. EST (Ottewaelder, et al.) Unpublished (1999) Contact: Ottewaelder B MPS			
TITLE	Am Klopferspitz 18a D-82152 Martinsried, Germany			
JOURNAL	This is the 5' sequence of the clone insert			
COMMENT	Clone from S. Wiemann, Molecular Genome Analysis, German Cancer Research Center (DKFZ). Email s.wiemann@dkfz-heidelberg.de; sequenced by Medigenomix (Martinsried/Germany) within the cDNA sequencing consortium of the German Genome Project..s1 sequence also available. This clone (DKFZp434F2018) is available at the RZPD in Berlin. Please contact the RZPD: Ressourcententrum, Heubnerweg 6, 14059 Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de. Location/Qualifiers 1..489 /organism="Homo sapiens" /db_xref="taxon:9606"			

```

/clone="DKFZp434F2018"
/clone_lib="434 (synonym: htes3)"
/tissue_type="testis"
/dev_stage="adult"
/lab_host="DH10B"
/notes="Vector: pSport1; Site_1: NotI; Site_2: SalI"

BASE COUNT      139 a      114 c      109 g      127 t
ORIGIN

Query Match      86.3%; Score 16.4; DB 9; Length 489;
Best Local Similarity 94.4%; Pred. No. 5.4e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 CTGGGGGCGAGTTATTGCA 18
        |||||||
DB      56 CTGGGGGCGAGTTATTGCA 39

RESULT 9
LOCUS      BI076955      524 bp      mRNA      linear      EST 20-JUN-2001
DEFINITION  NXPV_085.C02.F NXPV (NsF xylem Planings wood Vertical) Pinus taeda
ACCESSION   BI076955
VERSION     BI076955.1 GI:14515612
KEYWORDS    EST.
SOURCE      Loblolly pine.
ORGANISM    Pinus taeda
REFERENCE    Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
AUTHORS      Spermatophyta; Coniferopsida; Coniferales; Pinaceae; Pinus; Pinus.
TITLE        Sederoff, R.
JOURNAL      Molecular Basis of Wood Formation in the Pine Megagenome
COMMENT      Unpublished (2000)
              Contact: Johnson, Arthur
              North Carolina State University
              Tel: 919 515 7800
              Fax: 919 515 7801
              Email: ajohnson@unity.ncsu.edu
              Seq primer: 13.

FEATURES
    source
        1..524
            /organism="Pinus taeda"
            /strain="Coastal plain loblolly pine from North Carolina"
            /db_xref="taxon:3352"
            /clone="NXPV_085.C02"
            /clone_lib="NXPV (NsF xylem Planings wood Vertical)"
            /tissue_type="Xylem"
            /cell_type="Planings (secondary)"
            /dev_stage="Transitional"
            /lab_host="XLA-Blue"
            /note="Vector: Bluescript SK; Site_1: Eco RI; Site_2: XhoI
            ; The library is from early (spring) secondary wood, taken
            from a ten year old tree in the transitional phase. The
            tree is a kind gift of the Westvaco Corporation. Secondary
            xylem was harvested from the tree by peeling back the bark
            and primary xylem and then removing the underlying tissue
            with a block plane. NOTE: The sequences contain a 'cDNA
            adapter' between the EcoRI site and the start of the EST.
            The adapter sequence is 'AATTGGCAGAG'."

BASE COUNT      144 a      92 c      110 g      162 t
ORIGIN

Query Match      86.3%; Score 16.4; DB 13; Length 524;
Best Local Similarity 94.4%; Pred. No. 5.5e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 CTGGGGGCGAGTTATTGCA 18
        |||||||
DB      152 CTAGGGGCGAGTTATTGCA 169

RESULT 10

```

```

AM655709
LOCUS      AM655709      544 bp      mRNA      linear      EST 25-APR-2001
DEFINITION  106914 MARC 1BOV Bos taurus cDNA 5', mRNA sequence.
ACCESSION   AM655709
VERSION     AM655709.1 GI:7421535
KEYWORDS    EST.
SOURCE      cow.
ORGANISM    Bos taurus
REFERENCE    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS      Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
              Bovidae; Bovinae; Bos.
              1 (bases 1 to 544)
              Smith, R.P.L., Grosse, W.M., Freking, B.A., Roberts, A.J., Stone, R.T.,
              Casas, E., Wray, J.E., White, J., Cho, J., Fahrnenkrug, S.C., Bennett
              G.L., Heaton, M.P., Laegreid, W.W., Rohrer, G.A., Chitko-McKown, C.G.,
              Perce, G., Holt, T., Karaymicheva, S., Liang, F., Quackenbush, J. and
              Keeler, J.W.
              Sequence evaluation of four pooled-tissue normalized bovine cDNA
              libraries and construction of a gene index for cattle
              Genome Res. 11 (4), 626-630 (2001)
              21180013
              Contact: Smith TPL
              USDA, ARS, US Meat Animal Research Center
              PO Box 166, Clay Center, NE 68933-0166, USA
              Tel: 402 762 4366
              Fax: 402 762 4390
              Email: smith@email.marc.usda.gov
              Single pass sequencing. Bases called and alt-trimmed with phred
              v0.980904.e. Vector identified by cross-match with the -mnscore 18
              and -mismatch 12 options.
              PCR Primers
              FORWARD: AGGAACAGCTATGACCAT
              BACKWARD: GTTTCACAGCAGCAGC
              Plate: 80 row: 1 column: 21
              Seq primer: AATTAGGTGACACTATAG.

FEATURES
    source
        1..544
            /organism="Bos taurus"
            /db_xref="taxon:9913"
            /clone_lib="MARC 1BOV"
            /tissue_type="pooled"
            /lab_host="DH10B"
            /note="Vector: PCMV SPORT6; Site_1: NotI; Site_2: SalI;
            library made from pooled tissue from lymph node, ovary,
            fat, hypothalamus, and pituitary."

BASE COUNT      136 a      142 c      142 g      124 t
ORIGIN

Query Match      86.3%; Score 16.4; DB 10; Length 544;
Best Local Similarity 94.4%; Pred. No. 5.5e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 CTGGGGGCGAGTTATTGCA 18
        |||||||
DB      507 CTGGGGGCGAGTTATTGCA 524

RESULT 11
LOCUS      BH428192      701 bp      DNA      linear      GSS 12-DEC-2001
DEFINITION  BOHPV85TF BOHP Brassica oleracea genomic clone BOHPV85, DNA
ACCESSION   BH428192
VERSION     BH428192.1 GI:17613920
KEYWORDS    GSS.
SOURCE      Brassica oleracea
ORGANISM    Brassica oleracea
REFERENCE    Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
AUTHORS      Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
              Rosidae; eurosids II; Brassicales; Brassicaceae; Brassica.
              1 (bases 1 to 701)
              Town, C.D., Van Aken, S., Uterback, T. and Fraser, C.M.
              Whole genome shotgun sequencing of Brassica oleracea
              TITLE

```

JOURNAL Unpublished (2001)
COMMENT Contact: Chris Town

TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA.
Tel: 301-838-3523
Fax: 301-838-0208
Email: cdtown@tigr.org

DNA is from a doubled haploid provided by Tom Osborn.
Seq primer: TF

Class: sheared ends.

FEATURES
source Location/Qualifiers

1..701
/organism="Brassica oleracea"
/strain="TO100D03"
/db_xref="taxon:3712"
/clone="BOHPV85"
/clone.lib="BOHP"
/note="Vector: PHOS1, Site_1: BstXI, 2-3 kb sheared
genomic DNA inserted into PHOS1 using BstXI linkers"

BASE COUNT 211 a 136 c 104 g 250 t

Query Match 86.3%; Score 16.4; DB 17; Length 701;
Best Local Similarity 94.4%; Pred. No. 5.8e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 2 TGGGGGCGAGTTATTGCA 19
|||||
Db 301 TGGAGCGAGTTATTGCA 284

RESULT 12 AG132023 701 bp DNA linear GSS 04-NOV-2001
LOCUS Pan troglodytes DNA, clone: PTB-144E08.F, genomic survey sequence.
DEFINITION AG132023
ACCESSION AG132023
VERSION AG132023.1 GI:16661701

KEYWORDS GSS.
SOURCE Pan troglodytes male lymphoblast DNA, clone.lib:PTB Chimpanzee Male
BAC library clone:PTB-144E08.F.
ORGANISM Pan troglodytes

REFERENCE 1 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Pan.

AUTHORS Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
Totochi,Y., Watanabe,H. and Sakaki,Y.
TITLE BAC end sequences of library PTB

JOURNAL Unpublished
2 (bases 1 to 701)
Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
Totochi,Y., Watanabe,H. and Sakaki,Y.

TITLE Direct Submission
JOURNAL Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC):
1-7-22 Suenho-chou,Tsukumi-Ku, Yokohama, Kanagawa 230-0045, Japan
(E-mail:chimpsesgsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/
Tel:81-45-503-9111, Fax:81-45-503-9170)

COMMENT Clones are derived from the chimpanzee BAC library PTB This BAC end
was generated during the Rdp process and may have higher chance of
clone tracking errors.

PRIMERS
Sequencing: -21M13

LIBRARY
Vector : PKS145
R.site 1 : SacI
R.site 2 : SacI

FEATURES
source Location/Qualifiers

1..701
/organism="Pan troglodytes"
/db_xref="taxon:9598"
/clone="PTB-144E08.F"
/sex="male"
/cell_type="lymphoblast"

BASE COUNT 151 a 285 c 148 g 117 t
ORIGIN

Query Match 86.3%; Score 16.4; DB 17; Length 701;
Best Local Similarity 94.4%; Pred. No. 5.8e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 CTGGGGCGAGTTATTGCA 18
|||||
Db 517 CTGGGGCGAGTTATTGCA 500

RESULT 13 BJ153383 716 bp mRNA linear EST 24-JAN-2002
LOCUS BJ153383 Unpublished oligo-capped cDNA library, C. elegans L1 stage
DEFINITION Caenorhabditis elegans cDNA clone yk1318g06 3', mRNA sequence.
ACCESSION BJ153383
VERSION BJ153383.1 GI:18321368

KEYWORDS EST.
SOURCE Caenorhabditis elegans.
ORGANISM Caenorhabditis elegans

REFERENCE 1 Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidae
; Rhabditidae; Pelodermidae; Caenorhabditis.
AUTHORS Kohara,Y., Shin-I,T., Thierry-Mieg,J., Thierry-Mieg,D., Suzuki,Y.
and Sugano,S.
TITLE A complementary view of the C.elegans genome
JOURNAL Unpublished (2002)
COMMENT Contact: Tadasi Shin-I
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6855
Fax: 81-559-81-6855
Email: tsunagi@genes.nig.ac.jp.

FEATURES
source Location/Qualifiers

1..716
/organism="Caenorhabditis elegans"
/strain="N2"
/db_xref="taxon:6239"
/clone.lib="yk1318g06"
/clone.lib="unpublished oligo-capped cDNA library, C.
elegans L1 stage"
/sex="hermaphrodite"
/tissue.type="whole animal"
/dev.stage="L1"

/note="The AD-wrmcDNA library was generated with poly(A)+
RNA isolated from both hermaphrodite and male N2 worms of
all larval stages, embryos, adults and dauers and the
subsequent generation of cDNAs by poly(A) priming. The
cDNAs were cloned into pPC86"

BASE COUNT 202 a 131 c 163 g 220 t

ORIGIN
Query Match 86.3%; Score 16.4; DB 13; Length 716;
Best Local Similarity 94.4%; Pred. No. 5.8e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 CTGGGGCGAGTTATTGCA 18
|||||
Db 636 CTGGGGCGAGTTATTGCA 619

RESULT 14 BG282508 752 bp mRNA linear EST 21-FEB-2001
LOCUS BG282508 602406354F1 NIH_MGC_91 Homo sapiens cDNA clone IMAGE:4518661 5',
DEFINITION mRNA sequence.
ACCESSION BG282508
VERSION BG282508.1 GI:13031617
KEYWORDS EST.

SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
REFERENCE
1 (bases 1 to 752)
NIH-MGC http://mgc.ncl.nih.gov/
TITLE
National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL
Unpublished (1999)
COMMENT
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: DCTD/DRP
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM10413 row: k column: 14
High quality sequence stop: 486.
Location/Qualifiers
1..752
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4518661"
/clone_lib="NIH-MGC_91"
/tissue_type="adenocarcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: prostate; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
Average insert size 1.4 kb. Library enriched for
full-length clones and constructed by Life Technologies.
Note: This is a NIH-MGC Library."

BASE COUNT 232 a 225 c 153 g 142 t
ORIGIN

Query Match 86.3%; Score 16.4; DB 12; Length 752;
Best Local Similarity 94.4%; Pred. No. 5.9e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CTGGGGGCGAGTTATTCGA 18
|||||
DB 476 CTGGGGGCGAGTTATTCGA 459

RESULT 15
BH442712 866 bp DNA linear GSS 12-DEC-2001
LOCUS
DEFINITION BOGGG28TR BOGG Brassica oleracea genomic clone BOGGG28, DNA
sequence.
ACCESSION BH442712
VERSION BH442712.1 GI:17628426
KEYWORDS GSS.
SOURCE
ORGANISM Brassica oleracea.
Brassica oleracea.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Brassica.
1 (bases 1 to 866)
Town, C.D., Van Aken, S., Uterback, T. and Fraser, C.M.
Whole genome shotgun sequencing of Brassica oleracea
Unpublished (2001)
Other GSSs: BOGGG28TR
COMMENT
Contact: Chris Town
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA.
Tel: 301-838-3523
Fax: 301-838-0208
Email: cdtown@tigr.org
DNA is from a doubled haploid provided by Tom Osborn.
Seq primer: 7R
Class: sheared ends.
Location/Qualifiers
1..866
source

/organism="Brassica oleracea"
/strain="T01000DH3"
/db_xref="taxon:3712"
/clone="BOGGG28"
/clone_lib="BOGG"
/note="Vector: PHOS1; Site_1: BstXI; 2-3 kb sheared
genomic DNA inserted into PHOS1 using BstXI linkers"

BASE COUNT 295 a 144 c 182 g 245 t
ORIGIN

Query Match 86.3%; Score 16.4; DB 17; Length 866;
Best Local Similarity 94.4%; Pred. No. 6e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 TGGGGGCGAGTTATTCAC 19
|||||
DB 525 TGGAGGCGAGTTATTCAC 542

Search completed: July 1, 2003, 07:03:46
Job time : 864.385 secs

